2016

Computational Studies of Multi-Active Site Enzymes

Wanle Wei
University of Windsor

Computational Studies on Multi-Active Site Enzymes

By

Wanlei Wei

A Thesis

Submitted to the Faculty of Graduate Studies

through the Department of Chemistry and Biochemistry

in Partial Fulfillment of the Requirements for

the Degree of Master of Science

at the University of Windsor

Windsor, Ontario

2016

Computational Studies on Multi-Active Site Enzymes

by

Wanlei Wei

APPROVED BY:

Steven J. Rehse

Department of Physics

Bulent Mutus

Department of Chemistry & Biochemistry

James W. Gauld, Advisor

Department of Chemistry & Biochemistry

August 23, 2016

Declaration of Co-Authorship

I hereby declare that this thesis incorporates material that is a result of joint research as follows:

Chapter 3: was done in collaboration with Prof. Gerald Monard under the supervision of Prof. James W. Gauld.

Chapter 4: was done in collaboration with Prof. Gerald Monard under the supervision of Prof. James W. Gauld.

In both cases, key ideas, primary contributions, experimental designs, data analysis and interpretation, were performed by the author, and the contribution of co-authors was primary through the provision of guidance and computational resources

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis. I certify that the above material describes work completed during my registration as graduate student at the University of Windsor. I declare that, to the best of my knowledge, my dissertation does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my dissertation, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owners to include such materials in my dissertation. I

declare that this is a true copy of my dissertation, including any final revisions, as approved by my dissertation committee and the Graduate Studies office, and that this dissertation has not been submitted for a higher degree to any other University or Institution.

Abstract

Multi-scale computational approaches have been applied to investigate the catalytic mechanisms of (i) yeast mitochondrial threonyl-tRNA synthetase (MST1) pre-transfer editing and, (ii) glutamine deamination by glucosamine-6-phosphate synthase (GlmS).

MST1: MD and QM/MM-MD methods were used to examine (i) differences in the binding of its cognate and non-cognate Thr- and Ser-AMP substrates respectively, and (ii) mechanism of hydrolytic pre-transfer editing. In contrast to bound Thr-AMP, bound Ser-AMP is less constrained; i.e., greater positional variability, and as a result more waters are able to permeate the active site. Mechanistically, Thr-AMP hydrolysis occurs in two steps via a tetrahedral oxyanion intermediate. For Ser-AMP, however, formation of the oxyanion proceeds via a metastable intermediate while the second step, cleavage of the C_{carb}-OP bond, occurs as for Thr-AMP with similar energy barriers. Umbrella sampling shows that mechanism differences are due to a greater number of active site waters stabilizing the forming oxyanion in Ser-AMP, compared to Thr-AMP. As a result, the relative free energies of the rate-limiting barriers as well as that of the hydrolyzed products for Thr-AMP (14-19 and 4-10 kcal mol^{-1}, respectively) are markedly higher than for Ser-AMP (7-12 and 0-5 kcal mol^{-1}, respectively). That is, MST1 thermodynamically and kinetically preferentially edits against non-cognate substrate Ser-AMP, in agreement with experiment.

GlmS: MD and QM/MM studies were performed to examine the (i) protonation state of the mechanistically important amine of its N-terminal cysteinyl (Cys1) and its effect on its glutaminase domain and, (ii) mechanism by which it deaminates its glutamine substrate. Proton affinity studies suggest that at physiological pH, the Cys1-NH_2 group prefers to be neutral, and that if protonated, the active site is structurally less consistent. When the Cys1-NH_2 group acts as the required mechanistic base the rate limiting step

corresponds to nucleophilic attack of a water on the covalently cross-linked thioester intermediate with a free energy barrier of 78.2 kJ mol^{-1}.

Dedication

I dedicate this work to my family.

Acknowledgements

Throughout the last two years, there were many accounts of cheerful and rewarding memories. During this time, I worked with some amazing people and made long-lasting friendships. It made work at the office pleasant; for this, I am extremely grateful.

In particular, I would like to thank my supervisor, Prof James W. Gauld, for creating many opportunities, giving guidance, and proofreading during my Master's studies. It is due to the influence of Prof Gauld that I am now planning to pursue my doctoral degree in the near future.

I would like to thank my collaborating supervisor, Prof Gerald Monard, for his hosting me at the Université de Lorraine. Without his expertise in Computational Chemistry, I would not have learned nearly as much as I have during those 5 months; your insights have proven very valuable.

I would like to extend my appreciation and thanks to my committee, Prof Bulent Mutus and Prof Steven J. Rehse, for taking the time to read this thesis. In addition, it was a pleasure collaborating with Dr Mutus and his group members on several projects, which expanded my scientific horizon and curiosity.

The past and present members of the Gauld Group have each made the environment of the office brighter and more pleasant. I am thankful to Dr. Bogdan Ion, Daniel Simard, Mohamed Aboelnga, Travis DeWolfe, Sahar Nikoo, Sameer Jafar, and Taqred Al Nakhil for their helpful project related advices and suggestions, as well as countless entertainments. Additionally, I want to thank all of the undergraduate students: Claudia Lutfallah, Danielle Cooper, Jacqueline Gemus, Philip Stamov, Arielle Desamito, Sarah Henshaw, Paul Meister, Daniel Meister, and Julia Petta for their help along the way.

Last but not least, this thesis would not be complete without thanking those who offered kindness and support (and/or translation) during my short stay in Nancy: Prof

Antonio Monari, Prof Xavier Assfeld, Prof. Mounir Tarek, Prof Gerald Monard, Benjamin Meyer, Hugo Gattuso, Marilyne Viano, Marco Marazzi, Hatice Göken, Yiğitcan Eken, Taylan Turan, Li Sijin, Maurina Robiati, Guo Rui, Ruixue Liu, and Zhiwei Zhang. Each person mentioned above has made a positive, long-lasting impression upon me; I hope the very best in their future endeavors.

Table of Contents

List of Figures

List of Schemes

List of Tables

List of Abbreviations and Symbols

Abbreviations

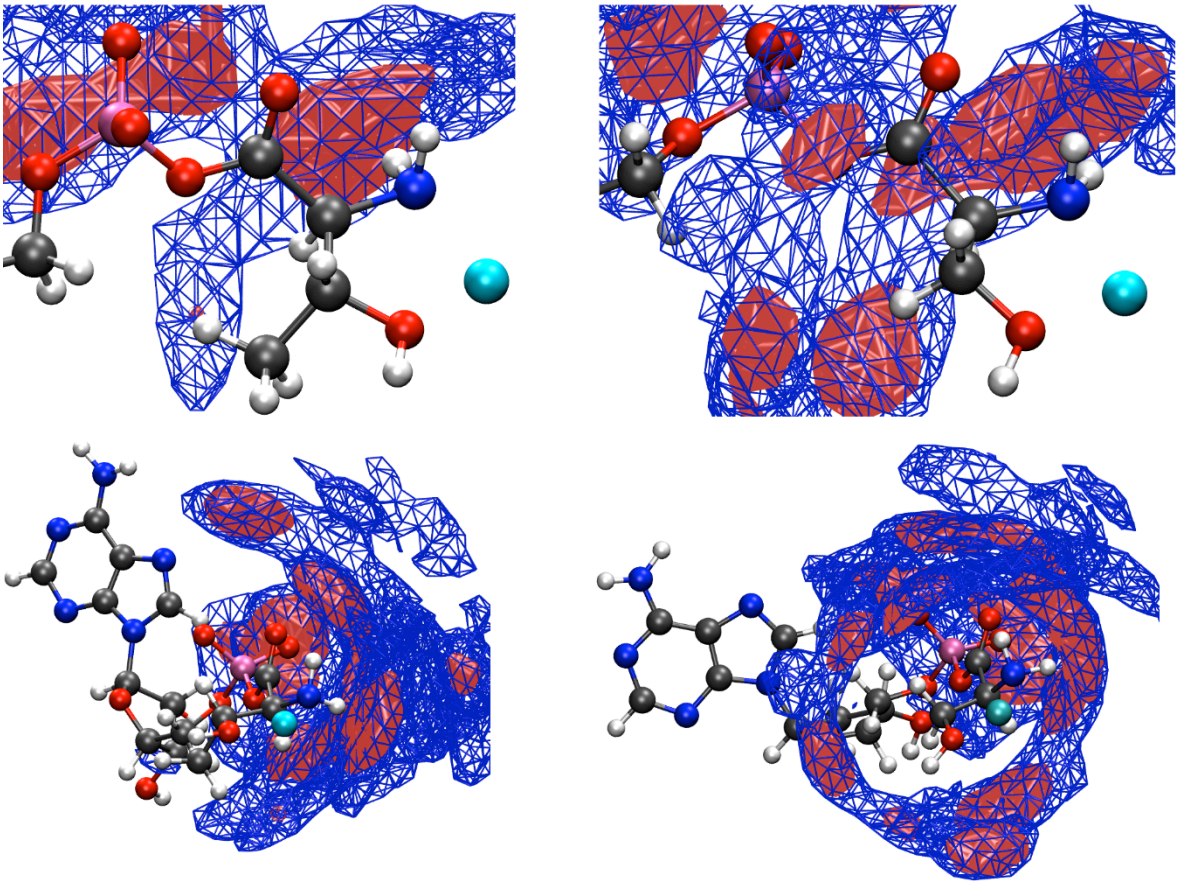
aa-AMP	Aminoacyl-adenylate
aaRS	Aminoacyl-tRNA synthetase
AMBER	Assisted Model Building with Energy Refinement (force field and/or MD program)
ATP	Adenosine triphosphate
CHARMM	Chemistry at HARvard Macromolecular Mechanics (force field)
DFT	Density functional theory
EE	Electronic Embedding
FF	Force Field
(R)ESP	(Restrained) Electrostatic potential
GlmS	Glucosamine-6-phosphate synthase
GlcN-6-P	Glucosamine-6-phoshpate
Gn-AT	Glutamine-dependent amidotransferase
IC	Intermediate Complex
LeuRS	Leucyl-tRNA synthetase
MD	Molecular Dynamics
ME	Mechanical Embedding
MM	Molecular Mechanics
MOE	Molecular Operating Environment (program)
MST1	Yeast mitochondrial threonyl-tRNA synthetase

ONIOM	Our own n-layered integrated molecular orbital and molecular mechanics
PA	Proton affinity
PC	Product complex
PES	Potential energy surface
PheRS	Phenylalanyl-tRNA synthetase
PP_i	Pyrophosphate
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
QM/MM-MD	Quantum Mechanics/Molecular Mechanics-Molecular Dynamics
RC	Reactant Complex
RMSD	Root-mean square deviation
Ser-AMP	Seryl adenylate
SP	Single point calculation
tRNA	Transfer ribonucleic acid
TS	Transition state
Thr-AMP	Threonyl adenylate
UDP-N-acetylglucosamine	Uridine diphosphate N-acetylglucosamine
WHAM	Weighted Histogram Analysis Method

Units symbols

Å	Angstroms
kcal mol^{-1}	Kilocalories per mole
kJ mol^{-1}	Kilojoules per mole

Chapter 1: Introduction to Computational Studies of Multi-Active Site Enzymes



1.1 Introduction

Enzymes are amino acid-based biological catalysts that are critical for carrying out life-essential reactions in all living systems. Consequently, over the decades they have been the focus of numerous studies into their structure, functions, inhibition etc. In particular, naturally, many of these studies have examined their catalytic mechanisms and the origin of their rate-enhancing power.[1,2] This has resulted in a range of effects being proposed as the sources of their catalytic abilities including, but not limited to: 1) proximity effects, in which the enzyme places the reactants in a favourable conformation for reaction,[3] 2) transition state stabilization, in which active site residues stabilizes the rate limiting step of the reaction,[3,4] and 3) alternative reaction pathway, in which a different, lower energy path is taken to arrive at the product from the reactant.[3]

Conventionally, enzymes were viewed as having a single active site and that such sites catalyze a single specific reaction. However, with the advent of new techniques in enzymology, our understanding of their roles and functions has been gradually evolving. For instance, it is now known that some enzymes are able to catalyze several different reactions within a single active site while others do not follow the one enzyme-one active site principle. In fact, several enzymes that harbour more than one catalytic domain can facilitate several different reactions,[5-8]

In the latter enzymes the active sites may be in constant communication with each other. In this work, communication is defined as the direct or indirect interaction between active sites. For example, the products of one domain may be passed onto another catalytic domain on the same enzyme where it then can undergo reaction.[9-11] This could be evolutionarily advantageous by increasing the catalytic efficiency of the enzyme without the need for further protein recruitment.

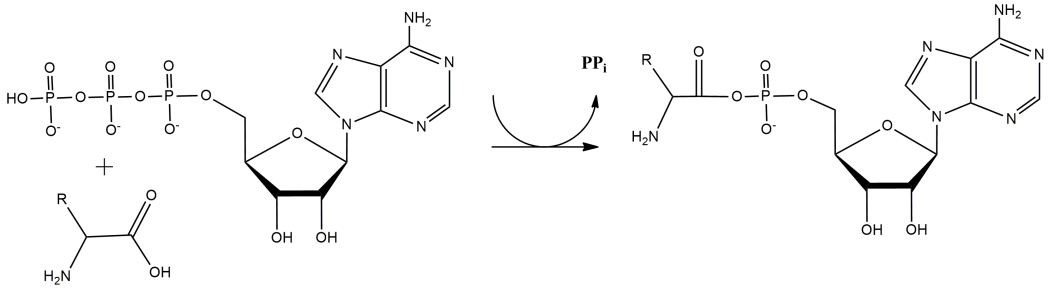
1.2 Aminoacyl-tRNA Synthetases

Amongst multi-active site enzymes, however, there are differences depending on the purpose(s) of the secondary active site.[9-11] For example, in some they may catalyse a reaction quite distinct from that performed by the first active site. Alternatively, it may partake in proofreading and cleavage of incorrect products from the primary active site. This is the case for many aminoacyl-tRNA Synthetases (aaRS), an ancient class of enzymes, where many members boast both an acylation domain and an editing domain.[11] Remarkably, aaRS are able to achieve a stated error rate of 1 in 10 000 and consequently, have been referred to as "paradigms" of molecular specificity.[12] The acylation domain is responsible for charging the aaRSs cognate tRNA with the correct amino acids, and utilizes the two half reactions shown in Scheme 1.1 This active site may also facilitate pre-transfer editing (Scheme 1.2a). For example, studies of isoleucyl-tRNA synthetases showed that its acylation site could hydrolyze valine, leading to the first reported discovery of pre-transfer editing.[13] Likewise, other amino acids (ex. methionyl tRNA synthetases) are also able to discern against chemically similar, but non-cognate amino acids such as the toxic homocysteine and homoserine.[14]

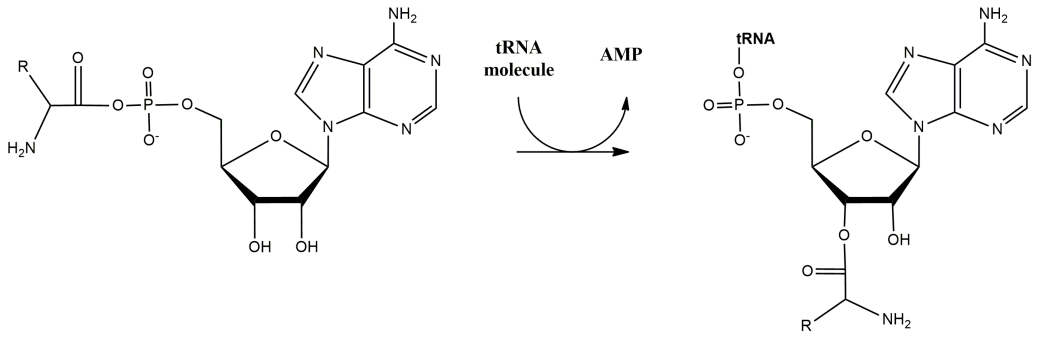
In contrast, the editing domain is responsible for cleavage of non-cognate aminoacyl moieties from the tRNA, also known as post-transfer editing (Scheme 1.2b) in cases where the first active site has mischarged the tRNA.[15]

Scheme 1.1 The two half-reactions catalysed by aminoacyl-tRNA synthetases in charging their cognate tRNA: (a) activation, (b) acylation.

(a)



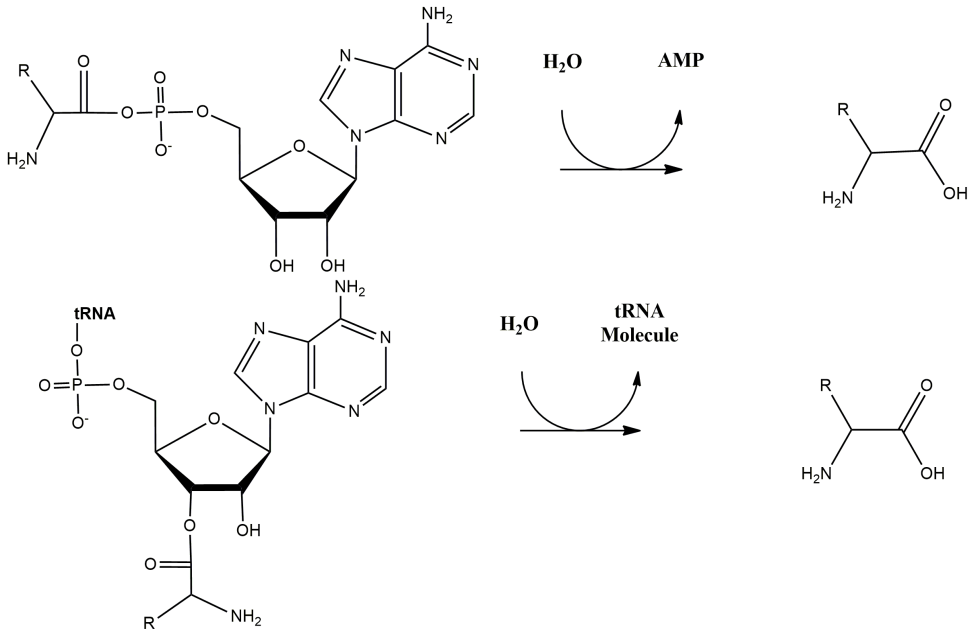
(b)



Scheme 1.2. The two known hydrolysis mechanisms of threonyl-tRNA synthetase: (a) pre-transfer editing, (b) post-transfer editing.

(a)

(b)



The roles of the editing domains of aaRSs have been experimentally and computationally studied.[12-14,16,17] For example, it has been experimentally shown that the mammalian threonyl tRNA synthetases (ThrRS) are able to discern between the cognate threonine and non-cognate serine amino acids.[11] This is remarkable given that these amino acids differ only by a methyl (CH_3) in their side-chains. More recently, Yeast Mitochondrial Threonyl-tRNA (MST1) has gained attention as a model enzyme to study pre-transfer editing due to its lack of a distal editing domain.[15] Experimental evidence

from AMP-formation assays have suggested that the hydrolysis of aminoacyl-AMP is possible without the presence of tRNA in the acylation domain.[12]

1.3 Glucosamine-6-phosphate Synthase

In contrast to aaRS, where the secondary active site functions as an editing domain, many multi-active site enzymes utilize their primary and secondary active sites synergistically to catalyze the final product[5-8]. For example, a subclass of glutamine-dependent amidotransferase (Gn-AT) enzymes, called Class II amidotransferases, share a common glutaminase active site.[9] More specifically, they are proposed to utilize a catalytic N-terminal cysteinyl (Cys1) to hydrolyze and remove ammonia from glutamine, releasing glutamic acid as a by-product (Equation 1.1). It is noted that the activation of the cysteinyl thiol group by formation of a thiolate has been mainly attributed to the terminal amine (NH_2) of the Cys1 residue.[9]

$$Glutamine + H_2O \longrightarrow Glutamic\ Acid + NH_3 \qquad \text{(Eq. 1.1)}$$

$$Acceptor + NH_3 \longrightarrow Product \qquad \text{(Eq. 1.2)}$$

The ammonia product from the first active site is then shuttled from the glutaminase domain to a second active site, through a water-sealed ammonia channel, more than 20 Å away.[18] It has been noted that most enzymes in this class can also use free exogenous ammonia.[9] The ammonia is subsequently used to create a variety of products at the synthase site depending on the identity of the actual enzyme (Equation 1.2).[9]

This class of enzymes is particularly important as they play key roles in a number of important metabolic reactions. For example, inhibition of glucosamine-6-phosphate synthase (GlmS) was found to alleviate diabetes in humans.[19] In addition, its crucial role in other species are evident as the deletion of GlmS was lethal in both fungi and insects.[19] This is not surprising as GlmS synthesizes the precursor to UDP-N-acetylglucosamine,

which itself is the precursor to a number of crucial products, such as: peptidoglycan, lipopolysaccharide (LPS), chitin, and other glycoproteins.[19] Other Gn-ATs also play a central role in the synthesis of key biomolecules such as purines (Glutamine PRPP Amidotransferase), pyrimidine (Carbomyl Phosphate Synthetase), asparagine (Asparagine Synthetases), and glutamate (Glutamate Synthetase), among many others.[9]

Despite the prevalence of multi-functional and/or multi-active site enzymes in biochemistry, the mode of information exchange between two or more active sites remains unclear. The complete understanding of the enzyme can only be achieved if the mechanistic details of each individual domains are elucidated. In this thesis, computational modelling approaches have been applied to investigate and attempt to shed light upon two exemplar multi-active site proteins – aaRS and Gn-AT.

1.4 References

(1) Liu, H. N.; Gauld, J. W. Substrate-assisted Catalysis in the Aminoacyl Transfer Mechanism of Histidyl-tRNA Synthetase: A Density Functional Theory Study. *J. Phys. Chem. B* **2008**, *112*, 16874-16882.

(2) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational enzymatic catalysis - clarifying enzymatic mechanisms with the help of computers. *PCCP* **2012**, *14*, 12431-12441.

(3) Bugg, T. D. H.: *Introduction to Enzyme and Coenzume Chemstry, 2nd Edition*; Blackwell Publishing, 2004.

(4) Schramm, V. L. Transition States and Transition State Analogue Interactions with Enzymes. *Acc. Chem. Res.* **2015**, *48*, 1032-1039.

(5) Wang, N.; McCammon, J. A. Substrate channeling between the human dihydrofolate reductase and thymidylate synthase. *Protein Sci.* **2016**, *25*, 79-86.

(6) Axelrod, H. L.; McMullan, D.; Krishna, S. S.; Miller, M. D.; Elsliger, M. A.; Abdubek, P.; Ambing, E.; Astakhova, T.; Carlton, D.; Chiu, H. J.; Clayton, T.; Duan, L.;

Feuerhelm, J.; Grzechnik, S. K.; Hale, J.; Han, G. W.; Haugen, J.; Jaroszewski, L.; Jin, K. K.; Klock, H. E.; Knuth, M. W.; Koesema, E.; Morse, A. T.; Nigoghossian, E.; Okach, L.; Oommachen, S.; Paulsen, J.; Quijano, K.; Reyes, R.; Rife, C. L.; van den Bedem, H.; Weekes, D.; White, A.; Wolf, G.; Xu, Q. P.; Hodgson, K. O.; Wooley, J.; Deacon, A. M.; Godzik, A.; Lesley, S. A.; Wilson, I. A. Crystal structure of AICAR transformylase IMP cyclohydrolase (TM1249) from Thermotoga maritima at 1.88 angstrom resolution. *Proteins-Structure Function and Bioinformatics* **2008**, *71*, 1042-1049.

(7) Huang, Y. M. M.; You, W. L.; Caulkins, B. G.; Dunn, M. F.; Mueller, L. J.; Chang, C. E. A. Protonation states and catalysis: Molecular dynamics studies of intermediates in tryptophan synthase. *Protein Sci.* **2016**, *25*, 166-183.

(8) Hasemann, C. A.; Istvan, E. S.; Uyeda, K.; Deisenhofer, J. The crystal structure of the bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase reveals distinct domain homologies. *Structure* **1996**, *4*, 1017-1029.

(9) Massiere, F.; Badet-Denisot, M. A. The mechanism of glutamine-dependent amidotransferases. *Cell Mol Life Sci* **1998**, *54*, 205-222.

(10) Sankaranarayanan, R.; Dock-Bregeon, A. C.; Romby, P.; Caillet, J.; Springer, M.; Rees, B.; Ehresmann, C.; Ehresmann, B.; Moras, D. The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **1999**, *97*, 371-381.

(11) Bushnell, E. A. C.; Huang, W. J.; Llano, J.; Gauld, J. W. Molecular Dynamics Investigation into Substrate Binding and Identity of the Catalytic Base in the Mechanism of Threonyl-tRNA Synthetase. *J. Phys. Chem. B* **2012**, *116*, 5205-5212.

(12) Zhou, X. L.; Ruan, Z. R.; Wang, M.; Fang, Z. P.; Wang, Y.; Chen, Y.; Liu, R. J.; Eriani, G.; Wang, E. D. A minimalist mitochondrial threonyl-tRNA synthetase exhibits tRNA-isoacceptor specificity during proofreading. *Nucleic Acids Res.* **2014**, *42*, 13873-13886.

(13) Fersht, A. R. Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. *Biochemistry* **1977**, *16*, 1025-1030.

(14) Fortowsky, G. B.; Simard, D. J.; Aboelnga, M. M.; Gauld, J. W. Substrate-Assisted and Enzymatic Pretransfer Editing of Nonstandard Amino Acids by Methionyl-tRNA Synthetase. *Biochemistry* **2015**, *54*, 5757-5765.

(15) Ling, J. Q.; Peterson, K. M.; Simonovic, I.; Soll, D.; Simonovic, M. The Mechanism of Pre-transfer Editing in Yeast Mitochondrial Threonyl-tRNA Synthetase. *J. Biol. Chem.* **2012**, *287*, 28518-28525.

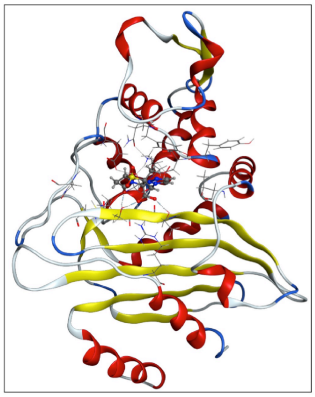
(16) Hagiwara, Y.; Nureki, O.; Tateno, M. Structural modelling of the complex of leucyl-tRNA synthetase and mis-aminoacylated tRNA(Leu). *FEBS Lett.* **2009**, *583*, 825-830.

(17) Hagiwara, Y.; Field, M. J.; Nureki, O.; Tateno, M. Editing mechanism of aminoacyl-tRNA synthetases operates by a hybrid ribozyme/protein catalyst. *J. Am. Chem. Soc.* **2010**, *132*, 2751-2758.

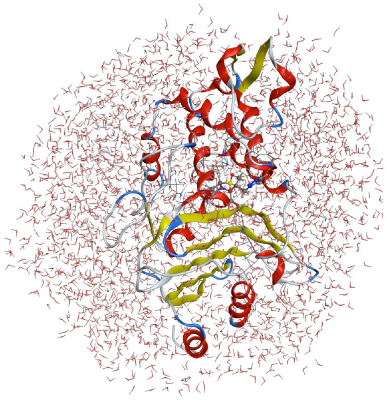
(18) Floquet, N.; Mouilleron, S.; Daher, R.; Maigret, B.; Badet, B.; Badet-Denisot, M. A. Ammonia channeling in bacterial glucosamine-6-phosphate synthase (Glms): molecular dynamics simulations and kinetic studies of protein mutants. *FEBS Lett.* **2007**, *581*, 2981-2987.

(19) Mouilleron, S.; Badet-Denisot, M. A.; Golinelli-Pimpaneau, B. Glutamine binding opens the ammonia channel and activates glucosamine-6P synthase. *J Biol Chem* **2006**, *281*, 4404-4412.

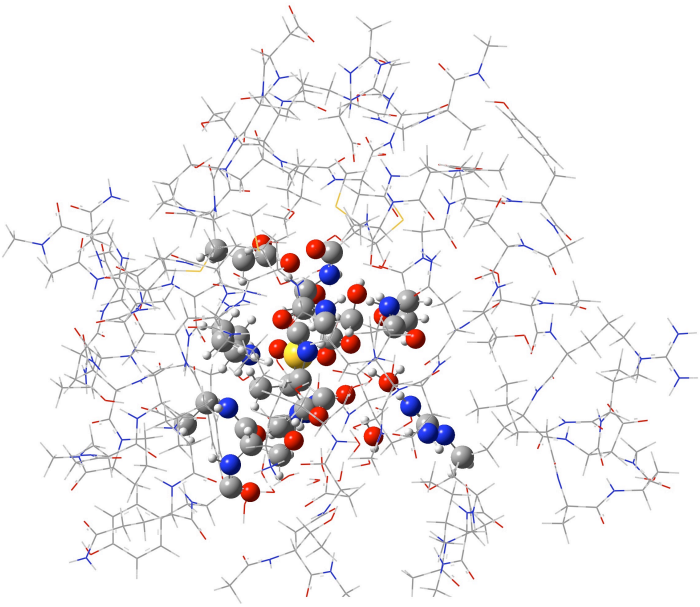
Chapter 2: Overview of Computational Methods



X-ray Crystal Structure



Molecular Dynamics



QM/MM

2.1 Introduction

Computational Chemistry is the use of computers to apply the theory of quantum chemistry to the study of chemical problems. This emerging field is crucial for understanding and rationalizing interactions at the atomic-scale, which is not observable by the naked eye. Computational studies often complement and give insights to experimental results. In the pharmaceutical industry for example, rational drug design processes are often guided initially by computationally scanning thousands of drug candidates.[1] In the later stages, experimental methods could synthesize those drugs which show promise.

In this chapter, several computational methodologies as applicable to enzymology will be introduced. For a clear understanding, it is a requirement to understand Schrodinger's Equation.

2.2 Schrodinger's Equation

Quantum Mechanics (QM) is a branch of physics that govern the behavior of subatomic particles including position, momentum, and energy. Many of these properties could be extracted from the time-dependent non-relativistic wavefunction (Eq. 2.1).[2]

$$\widehat{H}\psi(r,t) = i\hbar\frac{d\psi}{dt} \qquad \text{Eq. 2.1}$$

In the above equation, $\widehat{H}$ is the Hamiltonian energy operator and $\psi(r,t)$ defines the position of the subatomic particles. The time-dependent wavefunction is important when studying systems in which electron excitation occurs (ex. photon emission, radiative damage of DNA, photosynthesis).[3,4] In most cases however, the time-independent form is sufficient to accurately describe the behavior of molecular systems (Eq. 2.2).[4] The eigenvalue of the Hamiltonian operator (E), is the energy of the system.

$$\hat{H}\psi(r) = E\psi(r) \qquad \text{Eq. 2.2}$$

The complete Hamiltonian operator (Eq. 2.3) takes into account both the kinetic energy and potential energies of the nucleus and electrons.

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \frac{1}{2}\sum_{a=1}^{M}\frac{1}{\mathrm{M_a}}\nabla_a^2 + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} + \sum_{a=1}^{M}\sum_{b>a}^{M}\frac{Z_a Z_b}{r_{ab}} - \sum_{i=1}^{N}\sum_{a=1}^{M}\frac{Z_a}{r_{ia}} \qquad \text{Eq. 2.3}$$

The first two terms describe the kinetic energy contribution of the nuclei and electrons, respectively. The next three terms describes the potential energy generated by coulombic interactions between electrons, nuclei, and nucleus-electron, respectively. Unfortunately, Eq. 2.3 cannot be solved exactly but for the simplest of atoms. Consequently, approximations to the exact solution are required.[5]

Quantum calculations assume that electrons move in a field of positive nuclei due to the nucleus being three orders of magnitude larger than the mass of the electrons.[5] This is known as the Born-Oppenheimer approximation, which gives the electronic energy (Eq. 2.4). The nuclei-nuclei interaction energy is now a constant (V_{NN}) that could be added at the end of the calculation.[6]

$$\widehat{H_{el}} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} - \sum_{i=1}^{N}\sum_{a=1}^{M}\frac{Z_a}{r_{ia}} \qquad \text{Eq. 2.4}$$

Unfortunately, Eq. 2.4 is often impossible to be solve because the wavefunction of two or more electrons are inseparable since they interact through coulombic repulsion. The math for a three-body system is too complex, and a variety of approximations were devised (section 2.3).

2.3 Quantum Mechanical Description of Molecules

Several different quantum mechanical methods exist to model the reactivity and behavior of molecules, including: 1) Ab-initio approach, 2) Semi-Empirical Calculation (SE), and 3) Density Functional Theory (DFT).[7] These calculations are needed when accuracy is required, such as modelling kinetics, charge distributions, and transition states. Sole use of QM calculations on a system are called QM-Cluster calculations.[8]

2.4 Ab-initio Wavefunction Methods

The Hatree-Fock (HF) Method was one of the first to model electronic energy by breaking down the many-electron equation wavefunction into simpler single-electron problems.[7] It takes into account the electron-electron coulombic interactions by an average potential field generated by all other electrons, rather than explicit electron-electron repulsion terms. This is known as the central field approximation (CFA). CFA results in an erroneous assumption of HF, which is that the probability of finding an electron only depends on its distance from the nucleus, and not its distance from other electrons. It often leads to deviations from experimental values due to a lack of electron correlation.[7]

Consequently, many later methods introduced approximations for treating electron correlations, which are collectively known as post-Hartree-Fock methods. In conjunction with HF, they are known as ab-initio methods. These include methods introduced in the early 1930s, including Configuration Interaction (CI), Møller-Plesset (MP2, MP3, etc.), and Coupled Cluster (CC).[7] Some of these methods, such as CI, are in theory, exact, but also suffer higher computational costs.

2.5 Semi-Empirical Wavefunction Methods

Semi-Empirical Methods (SE) offer a compromise between speed and accuracy. SE are significantly faster for calculations because they replace many expensive and difficult

integral calculations with values obtained from experimental, ab-initio, or DFT calculations. For this reason, they could sometimes be more accurate than HF approach. SE methods also ignore core electrons.[7] Many modern SE methods are based on Modified Neglect of Diatomic Overlap (MNDO) Method, which is based on the HF method.[9] Over the years, AMx, and PMx methods have gradually improved on the MNDO method.[10] The AM1 method, introduced by James Stewart, improved the accuracy of atomic repulsions and hydrogen bonding.[11] PM3 is almost exactly identical to AM1, except that it aimed to use a different set of parametrization data set without human biases.[12] PM6 built upon the PM3 by using a larger set of parametrization data, and attempted to improve accuracy by pairwise parameters in addition to element-specific parameters.[13] It has been shown to work well for phosphorus, molybdenum, and magnesium, and is recommended for studying phosphates.[14-16] DFTB is a SE method based on Density Functional Theory (DFT), which has had successful application to a number of biomolecules.[17] In conclusion, certain SE methods give better results with some systems, and choosing the most compatible method is a necessary part of any theoretical study.

2.6 Density Functional Theory Methods

Since its discovery, Density Functional Theory (DFT) has become the method of choice for QM calculation. DFT calculates chemical properties using electron density which offers accuracy comparable to MP2 at a fraction of the computational cost.[7] There are five main types of functionals for DFT calculations: 1) Local Density Approximation (LDA), 2) Generalized Gradient Approximations (GGA), and 3) Hybrid Functionals, 4) Meta-GGA, and 5) Meta-Hybrid Functionals. The differences between these functionals is in their approach to approximate the exchange-correlation.[18] Poor approximations gives rise to problems in treating dispersion and van der Waals Forces. The B3LYP functional is a GGA approach which popularized DFT.[19] It was found to reproduce experimental values closely, and is still the gold standard used today. Recently however, Donald

Trhular developed the M06-2X meta-hybrid functional.[20] M062-X was found to describe chemical kinetics and long-range interactions more accurately than B3LYP as confirmed through many independent studies.[21] Consequently, the search for better and more accurate methods and functionals is still an active part of chemical research.

2.7 Significance of Different Basis Sets

Basis sets are set of functions used to describe the shape of the orbitals of an atom. Orbital descriptions rarely use the computationally expensive, but accurate Slater type orbitals (STO): $e^{-\zeta r}$ because a linear combination of Gaussian type orbitals (GTO): $e^{-\zeta r^2}$ give very accurate results.[7] Most SE calculations have a predefined basis set, while ab initio and DFT allow the user to specify the basis set used. The shape of all orbitals take the following form[7] (Eq. 2.5):

$$\varphi = Y_{lm} \sum_i C_i \sum_j C_{ij} e^{-\zeta_{ij} r^2} \qquad \text{Eq. 2.5}$$

In the above equation, Y_{lm} describes the type of the orbital (s, p, d, f-orbitals), C_i is the molecular orbital coefficient, which must be optimized through an iterative process, C_{ij} and ζ_{ij} are basis sets read from a database of values. Additionally, C_i may be a weighted sum of 1 to 9 primitive Gaussian functions called a contraction, which offers flexibility in describing the orbitals. The corresponding basis sets are called segmented basis sets.[7]

There are a number of different basis set options such as minimal, double-zeta, triple-zeta, and split-valence double-zeta basis sets, which differ in the number of GTO contractions that are used to describe each orbital. For example, the simplest is the minimal basis set, having the form: STO-nG. This notation means that a single contraction of 'n' GTO functions are used to approximate the shape of an orbital (ex. STO-3G). This is in contrast to double-zeta and triple-zeta basis sets, which uses 2 and 3

GTO contractions, respectively to describe a single orbital. A split-valence double-zeta basis set uses a single GTO to describe the core electrons, while using 2 GTO contractions for valence orbitals, since valence orbitals are of greater interest, chemically. An example of a split-valence double-zeta basis set is the Pople basis set such as 6-31G**, which is the most widely used basis set. 6-31G** uses 6 GTO functions to describe the core orbitals while using two contractions of 3 GTO and 1 GTO for the valence orbitals.[22] The split-valence triple-zeta function, such as 6-311++G(d,p) or 6-311+G(2df,p) is frequently used for single point calculations.[23,24] Sometimes, polarization functions could be added. A single asterisk (*) or (d) means that d-orbitals are added to non-hydrogen atoms, while a two asterisks (**) or (d,p) means that p-orbital are also added to hydrogens. (2df) means that 2 sets of d-orbitals, along with a single set of f-orbital are also added. Additionally, diffuse functions, such as (+) and (++) could be added to better describe distant regions of the orbital for non-first row elements and all elements, respectively.[7]

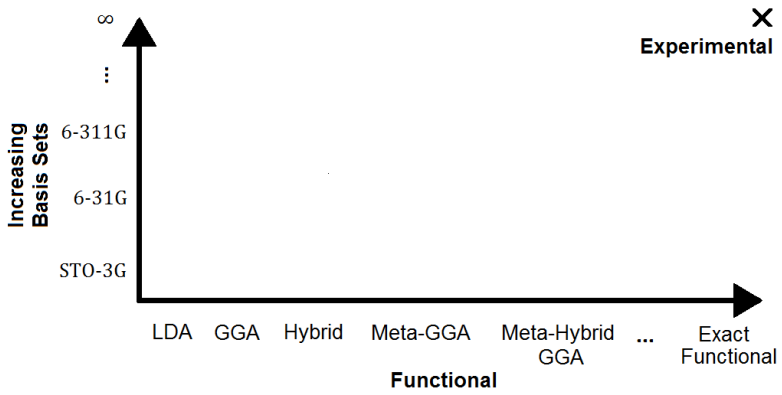


Figure 2.1. Pople's diagram comparing different basis sets and DFT methods to the accuracy of quantum calculations.

2.8 Molecular Mechanical Description of Molecules

Molecular Mechanical (MM) methods are modelled using the computationally less expensive Newtonian laws of motion. These differ from Quantum Mechanical (QM) methods, which are described by wavefunction or electron density (see section 2.3). MM

methods allow calculations to proceed significantly faster than QM methods, although reactivity and transition states cannot be modelled due to the lack of electron description.[25] MM describe each atom as a point charge of a certain mass. Bonds, angles, dihedral angles, van der Waals, and long range electrostatic interactions are modelled using simple equations of classical mechanics (Table 2.1). For example, bonds and angles are often represented as simple harmonic oscillators, with a spring constant (k) and an equilibrium distance (x_0 or Θ_0). The most common equations for describing MM are shown (Table 2.1).[25]

The Molecular Mechanics Force Field (MM FF) provides the constants in Table 2.1, which is molecule-dependent. For example, the force field parameters for the backbone atoms of valine and aspartic acid may differ significantly even if they share common atoms.[26] Currently, there are several types of FF in existence. Some are designed for the simulation of proteins, lipids, and nucleic acids such as AMBER, CHARMM, and GROMOS force fields,[27-29] while others are suitable for carbohydrates such as GLYCAM[30], and organo-transition metals such as OPLS.[31]

Table 2. 1. Common molecular mechanical force field (MM FF) equations and descriptors

Bond Distance	$E = \sum k_1(x - x_0)$	Harmonic Oscillator
Angle	$E = \sum k_2(\theta - \theta_0)$	Harmonic Oscillator
Dihedral Angle	$E = \sum A[1 + \cos(n\tau - \phi)]$	Periodic Function
Van der Waals Interactions	$E = \sum_i \sum_j -\frac{A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}}$	Lennard-Jones Potential
Electrostatic Interactions	$E = \sum_i \sum_j \frac{q_i q_j}{r_{ij}}$	Coulombic Energy

Although MM neglects the description of subatomic particles, it yields surprisingly accurate results during molecular dynamics simulations (MD). Many research groups

have successfully employed MD simulations for predicting various properties of proteins, such as radial distribution of water as well as protein dynamics.[32-34]

2.9 Quantum Mechanics/ Molecular Mechanics

Combined Quantum Mechanics/ Molecular Mechanics (QM/MM) became widely popular as a method of studying enzymatic catalysis.[35] Its main benefit was that it allowed the active site to be modelled using the highly accurate QM method, while the surrounding regions were described using MM. In this way, the conformation and charge effects of the low-layer (called 'real') onto the high-layer (called 'model') could be taken into account. There are generally two types of schemes when using QM/MM: additive and subtractive. [36] The additive scheme calculates the total energy ($E_{QM/MM}$) by combining the energies of high layer (E_{QM}) and the low layer (E_{MM}) with the boundary QM/MM boundary region (E_{QM-MM}) (Eq. 2.6). The latter is evaluated classically, but allows electrostatic charges at the boundary to propagate into the QM region.[37]

$$E_{QM/MM} = E_{QM} + E_{MM} + E_{QM-MM} \qquad \text{Eq. 2.6}$$

In contrast to the additive scheme, the subtractive scheme adds the energy of the high layer in QM to the energy of the entire system in MM. To avoid overcounting, the energy of the high layer in MM is removed. This allows the boundary region to be extrapolated, simplifying its treatment. This method is advantageous because it minimizes errors and artifacts arising from improper treatment of QM-MM interactions. The ONIOM methodology is one of the most used subtractive schemes developed by Morokuma and coworkers, which could be extended to n-number of layers.[37]

$$E_{QM/MM} = E_{QM,model} + E_{MM,real} - E_{MM,model} \qquad \text{Eq. 2.7}$$

2.10 Mechanical and Electrostatic Embedding

Mechanical embedding (ME) is applied by default in most QM/MM schemes, which describes the physical effects of bonded (stretching, bending, and torsional) interactions as well as non-bonded (van der Waals) interactions of the low layer and the high layer. This rigidity imposed on the high layer by the low layer helps keep the native conformation of the former. ME is evaluated in MM, and therefore requires that the high layer be defined in MM as well, which could require prior MM parameterization (section 2.8).[38] In many cases, ME has been sufficient in describing various chemical systems.[24]

Electrostatic embedding (EE), as the name suggests, allows electrostatics of the low layer to be taken into the QM Hamiltonian of the high layer. This method improves upon ME, and allows a more accurate description of the system.[38]

2.11 QM/MM Boundary

When the QM/MM boundary is present between a covalent bond, it creates unpaired electrons on each cut atom. This is an artifact. A common approach to rectify this problem is by placing a monovalent linker atom on the end of the bond.[38] The choice of the linker atom is arbitrary, but hydrogen atoms are most often used. One of the main problems with this method is that the linker atom does not represent the true atom which was replaced. To minimize this error, the researcher should chose the QM/MM boundary at a far enough distance from the reactive area so that the change in partial charges of the reaction does not propagate onto the linker atom.[39] This distance varies with different compounds, but is generally 3 bonds away from the location of bond breaking/ formation. To minimize errors, single bonds, non-polar bonds, and non-cyclic bonds should not fall on the QM/MM boundary.[40]

2.12 Explicit and Implicit Solvation

QM-cluster calculations are often done in gas-phase (vacuum) by default. In contrast, most biomolecules and enzymes are surrounded or accessible by water. As a consequence, solvation is usually required. There are two ways in which this could be achieved: Explicit and Implicit Solvation. In explicit solvation, water molecules or other solvents are added to the model of interest. This method is often used during molecular dynamics simulation in order to gain insight into detailed short-range interactions that cannot be seen otherwise (eg. hydrogen bonding interactions). However, it suffers disadvantages as it is computationally more expensive due to the added cost of modelling each water molecule. As a result, explicit solvation is generally only used during classical MD simulations.

In contrast, implicit solvation decreases computational time simply by generating a dielectric continuum around the solute, which is a field designed to mimic the properties of the solvent. There are a few different approaches of implicit solvation in the literature; the most two popular methods are: Polarizable Continuum Model (PCM) and Conductor-like Polarizable Continuum Model (CPCM).[41]

PCM gives good electrostatic energy results, although it discounts dispersion-repulsion energy and suffers a higher computational cost compared to other solvation methods.[42] It uses a number of overlapping spheres to map out a cavity held by the solute. The dipole of the solute is able to induce a reaction field in the adjacent solvent medium, which in turn could induce an electric field that interacts with the dipole. In general, the Hamiltonian of the time-independent PCM expression is given by Eq. 2.8, where $H°$ is the Hamiltonian in vacuum and V_{s-s} is the interaction of the solute with the solvent.[41]

$$H = H° + V_{s-s} \qquad \text{Eq. 2.8}$$

CPCM is a variation of the PCM, and treats the dielectric continuum as a conductor, which impacts the polarization between the solute and the solvent.[43] As a consequence, it more accurately models solvents with a higher permittivity and conductance. At the same time, it is computationally less expensive than the PCM method. Other methods do exist, but are less commonly used compared to PCM and CPCM.

2.13 Molecular Dynamics Simulation

Molecular Dynamics Simulations (MD) are describes the system as a progression of time *in silico*. They give information pertaining to structure, function, as well as key interactions such as hydrogen bonding, π- π stacking, and long-range interactions. MD is often used for studying the interaction of a protein with potential drug candidates. It is also used in the first step of a multi-scale computational study. There are 3 common types of MD simulations depending on the level of theory used: 1) Classical MD, 2) QM/MM-MD, and 3) Quantum Dynamics. Unfortunately, the latter is currently computationally too expensive to be applicable to biomolecules, and will not be further covered here. MD simulations have conventionally used classical mechanics, which is widely used for protein dynamics. With advances in computational algorithms and power, it is now possible to perform classical MD simulations on the order of microseconds.[34] Longer simulation allow improved sampling of biological molecules, which became possible only recently. Classical MD simulations are available in many scientific computer programmes such as AMBER, CHARMM, NAMD, and GROMACS.[27-29,44] Many of these programmes have also recently implemented QM/MM-MD for studying molecular systems. Like QM/MM (section 2.9), QM/MM-MD allows the QM region to be studied using a higher level of theory while keeping the rest of the protein in a lower level of theory. The QM layer allows chemical reactions to be studied dynamically in contrast to QM/MM, in a method called umbrella sampling (section 2.14). One limitation of umbrella sampling is that ab-initio wavefunctions and DFT are computationally too

expensive to be applied to QM/MM-MD.[34] Consequently, SE methods are used for the QM region, limiting its accuracy.

2.14 Choosing the Correct Molecular Dynamics Time step

The smallest observable changes in the system should dictate the time step of the MD simulation. For classical MD, this would correspond to the hydrogen bond stretching which occurs every 1 fs.[45] Employing the SHAKE algorithm constrains this bond, allowing the timestep to increase to 2 fs, which speeds up the calculation without destabilizing the simulation.[46] For QM/MM-MD, the time step must be further decreased to 0.5 fs to accommodate the faster observable atomic fluctuations due to treatment by QM.[34]

2.15 Treatment of Molecular Dynamics Boundaries

In silico simulations try to emulate the bulk solvent environment of soluble proteins. Two approaches exist when dealing with the boundary conditions of the MD simulation: Minimal Solvation (eg. Layered or spherical solvation) and Periodic Boundary Condition (PBC). The former saves computational time since it creates a layer or sphere of few water molecules surrounding the solute.[23,24,33] However, diffusion of the solute during longer simulations could cause it to interact with the edge of the solvent box, leading to non-bulk solvent conditions. This is an artifact. As a result, minimal solvation is often used for short simulations, such as equilibration of a biomolecule in preparation for a multi-scale computational study.[23,24,33]

Periodic Boundary Condition is the preferred method for longer simulations where diffusion plays an important role.[34] In PBC, solvation is often done as a rectangular prism or a truncated octahedron. This is because PBC creates infinite number of virtual images of the real image in the X,Y,Z-directions, and many shapes do not allow tight packing (eg. sphere). Counter ions (eg. often Na^+ and Cl^- ions) must be added to avoid summing to an

infinite charge due to its infinite periodicity.[47] This is because charged systems renders the equation unsolvable. Diffusion is allowed in PBC because virtual images mimic the behavior of the real image. As a result, atoms which diffuses out of the real image reappear on its opposite end. A large enough box is required in order to minimize the artifact arising from the interactions of the solute with its virtual image.[34] In general, PBC is computationally more expensive than layered or spherical solvation, but also more accurate.

2.16 Umbrella Sampling

"Umbrella sampling uses biased potentials along (one- or more-dimensional) reaction coordinates to drive a system from one thermodynamic state to another (eg. reactant, transition state, and product)".[48] A biased potential is required because certain regions of interest, such as transition states, occur infrequently due to its higher relative energy. The biased potential is applied discretely along the reaction path, and a probability distribution is created for each sampling window. Later, the weighted histogram analysis method (WHAM) is used to compile and relate the probability in each window to the overall free energy landscape (Figure 2.2).[49]

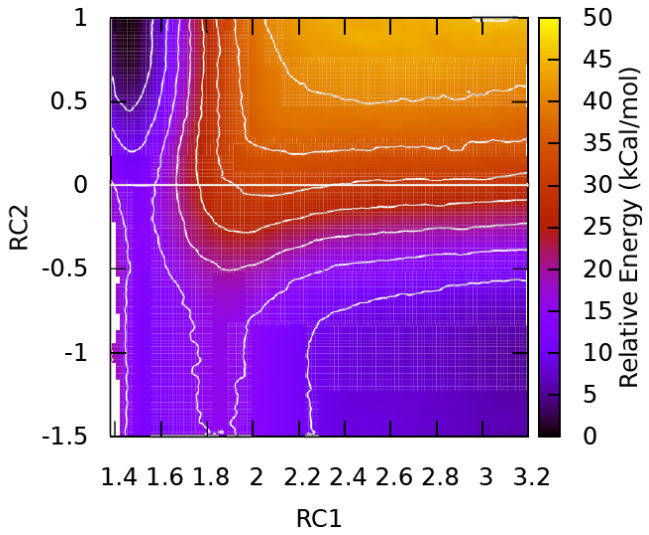


Figure 2.2. Example of a free energy surface (PES) obtained from umbrella sampling through QM/MM-MD. The above surface represents the hydrolysis of threonyl-AMP within the aminoacylation active site of threonyl-tRNA synthetase.

Umbrella sampling is applied to MD simulations so that chemical reactivity and classical properties could be studied dynamically. Another advantage of QM/MM-MD umbrella sampling is that relative free energy could be derived from the probability distribution of thermodynamic states, whereas traditional QM/MM optimizations rely on a few additional corrections (eg. free energy corrections and zero point energy corrections). Although QM/MM-MD is the recommended method for studying enzymatic catalysis, it suffers high computational costs when using ab-initio and DFT methods (section 2.13).[50] With the promise of new algorithms and hardware (eg. GPU-enabled QM/MM MD[27]) these calculations are expected to become more feasible in the near future.

2.17 References

(1) Lin, C. H.; Hsieh, Y. S.; Wu, Y. R.; Hsu, C. J.; Chen, H. C.; Huang, W. H.; Chang, K. H.; Hsieh-Li, H. M.; Su, M. T.; Sun, Y. C.; Lee, G. C.; Lee-Chen, G. J. Identifying

GSK-3 beta kinase inhibitors of Alzheimer's disease: Virtual screening, enzyme, and cell assays. *Eur. J. Pharm. Sci.* **2016**, *89*, 11-19.

(2) Scully, M. O. The Time Dependent Schrodinger Equation Revisited I: Quantum field and classical Hamilton-Jacobi routes to Schrodinger's wave equation. *395th Wilhelm and Else Heraeus Seminar - Time Dependent Phenomena in Quantum Mechanics* **2008**, *99*.

(3) Petersilka, M.; Gossmann, U. J.; Gross, E. K. U. Excitation energies from time-dependent density-functional theory. *Phys. Rev. Lett.* **1996**, *76*, 1212-1215.

(4) Zhekova, H.; Krykunov, M.; Autschbach, J.; Ziegler, T. Applications of Time Dependent and Time Independent Density Functional Theory to the First pi to pi* Transition in Cyanine Dyes. *J. Chem. Theory Comput.* **2014**, *10*, 3299-3307.

(5) Levine, I. N.: *Quantum Chemistry*: Prentice Hall: New Jersey, 1991.

(6) Szabo, A. O., N: *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications, 1996.

(7) Young, D.: *Computational Chemistry- A Practical Guide for Applying TEchnique to Real World Problems*; John Wiley & Sons, Inc: Canada, 2001.

(8) Sumner, S.; Soderhjelm, P.; Ryde, U. Effect of Geometry Optimizations on QM-Cluster and QM/MM Studies of Reaction Energies in Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4205-4214.

(9) Thiel, W. V., A. Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group. *J Phys. Chem* **1996**, 616-626.

(10) Stewart, J. J. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. *J. Mol. Model.* **2004**, *10*, 155-164.

(11) Dewar, M. Z., E; Healy, E; Stewart, J. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc* **1984**, 3902-3909.

(12) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209-220.

(13) Stewart, J. J. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173-1213.

(14) Imhof, P.; Noe, F.; Fischer, S.; Smith, J. C. AM1/d parameters for magnesium in metalloenzymes. *J. Chem. Theory Comput.* **2006**, *2*, 1050-1056.

(15) Voityuk, A. A.; Rosch, N. AM1/d parameters for molybdenum. *J. Phys. Chem. A* **2000**, *104*, 4089-4094.

(16) Lopez, X.; York, M. D. Parameterization of semiempirical methods to treat nucleophilic attacks to biological phosphates: AM1/d parameters for phosphorus. *Theor. Chem. Acc.* **2003**, *109*, 149-159.

(17) Zheng, G. S.; Irle, S.; Morokuma, K. Performance of the DFTB method in comparison to DFT and semiempirical methods for geometries and energies Of C-20-C-86 fullerene isomers. *Chem. Phys. Lett.* **2005**, *412*, 210-216.

(18) Peverati, R.; Truhlar, D. G. Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philos Trans A Math Phys Eng Sci* **2014**, *372*, 20120476.

(19) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988**, *38*, 3098-3100.

(20) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(21) Villuendas-Rey, Y.; Alvarez-Idaboy, J. R.; Galano, A. Assessing the Protective Activity of a Recently Discovered Phenolic Compound against Oxidative Stress Using Computational Chemistry. *Journal of Chemical Information and Modeling* **2015**, *55*, 2552-2561.

(22) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724-728.

(23) Fortowsky, G. B.; Simard, D. J.; Aboelnga, M. M.; Gauld, J. W. Substrate-Assisted and Enzymatic Pretransfer Editing of Nonstandard Amino Acids by Methionyl-tRNA Synthetase. *Biochemistry* **2015**, *54*, 5757-5765.

(24) Dokainish, H. M.; Gauld, J. W. Formation of a Stable Iminol Intermediate in the Redox Regulation Mechanism of Protein Tyrosine Phosphatase 1B (PTP1B). *ACS Catalysis* **2015**, *5*, 2195-2202.

(25) Jensen, F.: *Introduction to Computational Chemstry, 2nd Ed*; John Wiley & Sons, Ltd: West Sussex, 2007.

(26) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J. Am. Chem. Soc.* **1996**, *118*, 2309-2309.

(27) D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman. AMBER 2015. *University of California, San Francisco* **2015**.

(28) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain $\chi1$ and $\chi2$ Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257-3273.

(29) Reif, M. M.; Hünenberger, P. H.; Oostenbrink, C. New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *J. Chem. Theory Comput.* **2012**, *8*, 3705-3723.

(30) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A generalizable Biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622-655.

(31) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.

(32) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*.

(33) Dokainish, H. M.; Gauld, J. W. A Molecular Dynamics and Quantum Mechanics/Molecular Mechanics Study of the Catalytic Reductase Mechanism of Methionine Sulfoxide Reductase A: Formation and Reduction of a Sulfenic Acid. *Biochemistry* **2013**, *52*, 1814-1827.

(34) Ugur, I.; Marion, A.; Aviyente, V.; Monard, G. Why Does Asn71 Deamidate Faster Than Asn15 in the Enzyme Triosephosphate Isomerase? Answers from Microsecond Molecular Dynamics Simulation and QM/MM Free Energy Calculations. *Biochemistry* **2015**, *54*, 1429-1439.

(35) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227-249.

(36) van der Kamp, M. W.; Mulholland, A. J. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**, *52*, 2708-2728.

(37) Chung, L. W.; Sameera, W. M.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H. B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678-5796.

(38) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185-199.

(39) Shusuke, Y.; Koki, U.; Kazuto, N.; Mitsutaka, O.; Kizashi, Y.; Haruki, N. How to determine boundaries for QM/MM calculations: A guideline based on linear response function for glutathione. *Journal of Physics: Conference Series* **2013**, *454*, 012055.

(40) Matta, C. F.: *Quantum Biochemistry*; Wiley-VCH Verlag GmbH & Co KGaA: Germany, 2010.

(41) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999-3093.

(42) Implicit Solvent Models. July 31, 2016).

(43) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669-681.

(44) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

(45) Choe, J. I.; Kim, B. Determination of proper time step for molecular dynamics simulation. *Bull. Korean Chem. Soc.* **2000**, *21*, 419-424.

(46) Krautler, V.; Van Gunsteren, W. F.; Hunenberger, P. H. A fast SHAKE: Algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **2001**, *22*, 501-508.

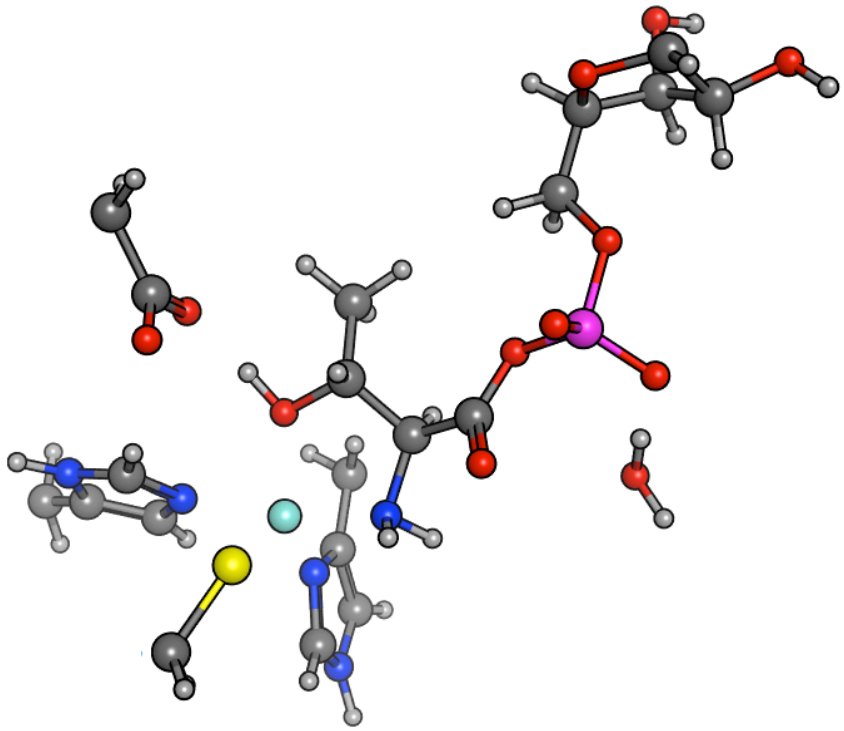
(47) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306-317.

(48) Kastner, J. Umbrella sampling. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2011**, *1*, 932-942.

(49) Grossfield, A. WHAM: The weighted histogram analysis method, version 2.0.9. *University of Rochester Medical Center, Rochester, NY*.

(50) Bushnell, E. A. C.; Berryman, V. E. J.; Gauld, J. W.; Boyd, R. J. The Importance of the MM Environment and the Selection of the QM Method in QM/MM Calculations: Applications to Enzymatic Reactions. *Combined Quantum Mechanical and Molecular Mechanical Modelling of Biomolecular Interactions* **2015**, *100*, 153-185.

Chapter 3: Pre-transfer Editing in Threonyl-tRNA Synthetase: Roles for Differential Solvent Accessibility and Intermediate Stabilization



3.1 Introduction

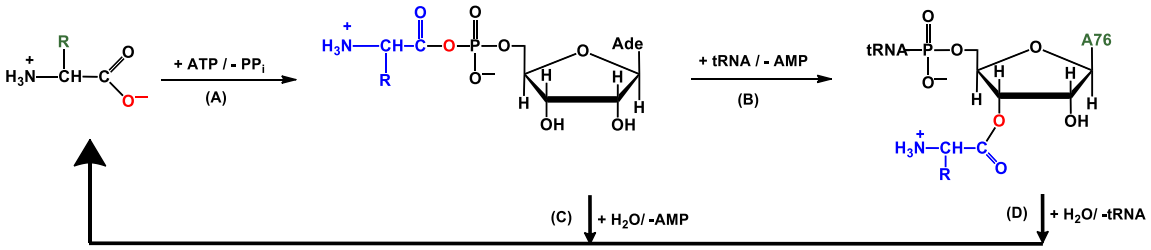
Accurate translation of genetically encoded material into proteins is a fundamental biological process critical for maintaining normal cellular functions.[1] A key step in this process is the attachment of amino acids to their cognate tRNA in preparation for incorporation into proteins. This step is catalyzed via two half-reactions by the aminoacyl-tRNA Synthetases (aaRSs)[2-4] class of enzymes which are also known to be important to a diverse array of physiological processes including viral assembly,[5,6] cancer,[7] and porphyrin biosynthesis.[8] For instance, ThrRS has been shown to be one possible clinical marker, along with other well-known cancer proteins (e.g. TNF-α), of human ovarian cancer through its role in angiogenesis.[7] In addition, it has also been shown to be the main culprit in myositis: an autoimmune disease associated with inflammation of muscles.[9,10]

In the first half-reaction of aaRSs the amino acid is reacted with ATP to give the aminoacyl-adenylate (aa-AMP) derivative (Scheme 3.1.a). Then, within the same active site they catalyze via a substrate-assisted mechanism the transfer of the aminoacyl moiety onto its cognate tRNA ($tRNA^{aa}$; Scheme 3.1.b).[11-13] However, the existence of near-cognate amino acids presents problems in cognate amino acid recognition. In fact the aminoacylation site of ten of the 24 aaRSs also misacylate their near-cognate amino acids,[14] potentially jeopardizing the process of protein synthesis. Remarkably, however, aaRSs achieve aminoacylation with an exceptionally high-fidelity, having a suggested intrinsic average error rate of about 1 in 10 000.[15]

There are several known possible approaches used by aaRS in order to achieve such exceptional accuracy. In particular, many possess proof-reading mechanisms which degrade either misactivated aa-AMP (ie. pre-transfer editing: Scheme 3.1.c) or misacylated aa-$tRNA^{aa}$ (ie. post-transfer editing: Scheme 3.1.d).[16,17] Post-translational editing sites, found in a distal region of several aaRS separate from the acylation active

site, have been studied quite extensively.[18-20] In contrast, pre-transfer editing which refers to the hydrolysis of aa-AMP in this paper, has been less widely studied,[21] and the detailed mechanisms by which pre-transfer editing occurs are currently unknown.[22] This is unfortunate given that defective editing mechanisms have been linked to a wide array of problems such as cellular growth defects, mitochondrial dysfunction, and neurodegeneration.[23]

Scheme 3.1. The two half-reactions catalysed by aminoacyl-tRNA synthetases in charging their cognate tRNA: (A) activation, and (B) acylation, and the two major editing pathways: (C) pre-transfer, and (D) post-transfer editing.



The occurrence of pre-transfer editing was first inferred from kinetics data, that suggested that in the case of IleRS, the accumulation of misaminoacylated Val-tRNA^{Ile} was smaller than expected.[24] However, it was not clear at the time if this process was a significant contributor to the overall physiological editing process.[14] More recently, a study of LeuRSs by Tan and coworkers[25] suggested that there are species-specific preferences for pre- and post-transfer editing.[25,26] For example, although editing mechanisms of LeuRS from *Aquifex aeolicus* and *Escherichia coli* had comparable overall efficiencies, the former preferred pre-transfer editing against near-cognate amino acids while the latter exhibited a great propensity for post-transfer editing.[25] A similar preference of one editing approach over another was also observed for PheRSs found in eukaryotes, archaea, and bacteria.[26] More specifically, the former two domains (eukaryotes and archaea) are suggested to predominantly utilize pre-transfer while the

latter, at least in *E. coli*, has a bias for post-transfer editing. These preferences of one editing approach over another may help provide insight into how some aaRSs that lack a distal editing site are still able to charge their cognate tRNA with high fidelity.

For instance, ThrRS and ProRS from *E. coli* cannot accurately discriminate between cognate and near-cognate amino acids solely based on pre-transfer editing.[26,27] In contrast, ThrRS from yeast mitochondria (MST1) lacks the distal editing domain usually found in other ThrRS.[28] Yet, in an experimental study by Ling and coworkers,[28] they showed that in MST1 pre-transfer editing selectively hydrolyzed 80% of the sterically smaller, but chemically similar near-cognate substrate seryl-AMP compared to 20% of the cognate substrate threonyl-AMP. They also concluded that pre-transfer editing occurs in the acylation domain of MST1 in the absence of its corresponding tRNA ($tRNA^{Thr}$). More recently, however, Zhou and coworkers[15] showed that pre-transfer editing could also occur in the same active site when $tRNA^{Thr}$ is also bound, but at an enhanced rate. Kinetic and mutational studies have also suggested that hydrolytic pre-transfer editing has a larger role in ensuring correct product fidelity when the rate of aminoacylation is lower.[29] Unfortunately, however, the exact details of how MST1 achieves pre-transfer editing within the acylation active site remains unknown. For instance, despite only differing by a $-CH_2-$, it has been suggested that the binding of Thr-AMP versus Ser-AMP may affect the position of nearby waters that could play a role in substrate discrimination.[13]

We have performed microsecond MD simulations to compare the active site structures, water densities, and interactions of unbound yeast mitochondrial ThrRS (MST1) with bound Thr-AMP or Ser-AMP. In addition, QM/MM-MD umbrella sampling has been used to elucidate the catalytic mechanism of pre-transfer editing in MST1, in order to explain its preference for hydrolysis of the non-cognate substrate Ser-AMP.

3.2. Computational Methods

3.2.1 Preparation of the Sample.

The X-ray crystal structure of a single monomer of yeast mitochondrial threonyl-tRNA synthetase (MST1) from *Saccharomyces cerevisiae* (PDB ID: 3UH0, 2.0 Å resolution) was used as the starting structure for the computational studies.[22] In silico mutations of the existing bound substrate analogue (threonyl sulfamoyl adenylate) to threonyl adenylate (Thr-AMP) and seryl adenylate (Ser-AMP) were manually performed.

The system was hydrogenated in accordance with the protonation states of various charged side chains with PROPKA,[30-33] and histidine residues were protonated according to their polar environment. The tleap module of AMBER14[34] was used to build topology and coordinate files. The system was solvated with TIP3P[35] water molecules, resulting in cubic boxes with an edge length of ~105.5 Å. The solvation achieved a density of 1.01 g/cm^3 after equilibration, and contained ~37000 water molecules (for a total of 120000 atoms). This corresponds to a protein concentration in the box of approximately one quarter of the original MST1 crystal structure.

3.2.2 Molecular Dynamics Simulations.

All simulations were performed using the AMBER14 program package[34] with the pmemd module, and the cuda-enabled graphics processing units (GPUs) version of pmemd.[36,37] Furthermore, all used the ff14SB[38] and TIP3P[35] force fields for protein and waters, respectively. The specific N-terminal seryl and threonyl fragments of the Ser-AMP and Thr-AMP ligands respectively, were built with the antechamber[39,40] program using the ff14SB force field and RESP atomic charges. It should be noted that the latter charges were derived from gas-phase optimized structures of seryl and threonyl fragments, capped by methyl groups, obtained at the HF/6-31G(d) level of theory using the Gaussian09[41] program. Periodic boundary conditions were applied, and an NVT ensemble

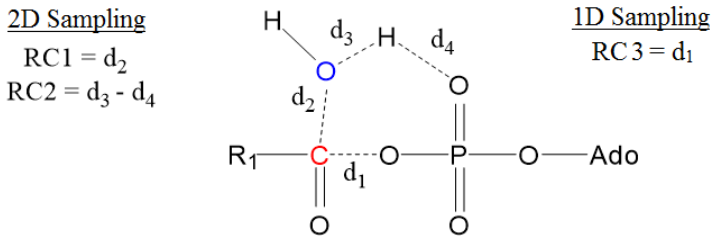
was used with Anderson temperature coupling scheme.[42] The Particle-Mesh Ewald (PME) procedure was used to calculate long-range electrostatic interactions and a cut-off of 8 Å was applied in real space. The SHAKE algorithm was used to restrain the bond stretches involving hydrogen atoms, thus allowing a 2 fs time step to be used for both equilibration and production. The equilibrations were conducted in five stages after energy minimization: (1) To ensure the proper geometry of the hydrogen atoms, all heavy atoms, including water oxygen, were restrained with a harmonic constant of 50 kcal mol^{-1} $Å^{-2}$ for 100 ps, at 10 K; (2) The same constant and conditions were applied for an additional 100 ps, but with the removal of the restraint on the water oxygens to ensure optimized positions of water with respect to the protein environment; (3) The restraint harmonic constant on the protein heavy atoms was decreased to 5 kcal mol^{-1} $Å^{-2}$ for 100 ps; (4) Subsequently, the harmonic constant was removed for 100 ps; (5) The system was gradually heated to 300 K over a time period of 2000 ps. The velocities were randomly updated every 10 steps for equilibration stages 1-4 and every 100 steps for stage 5. The subsequent production MD simulation was for a 1 μs period of time. Notably, they were performed with five harmonic restraints imposed in order to keep the penta-coordination[4] of the structurally important Zn^{2+} with thiolate sulfur of Cys96 (300 kcal mol^{-1} A^{-2}), ε nitrogen of His147 (50 kcal mol^{-1} A^{-2}), and δ nitrogen of His277 (150 kcal mol^{-1} A^{-2}). The two remaining restraints were either the amine (150 kcal mol^{-1} A^{-2}) and alcohol (200 kcal mol^{-1} A^{-2}) group of the ligand in the case of aminoacyl-AMP, or two waters (100 kcal mol^{-1} A^{-2}) in the case of the apoenzyme. The restraint harmonic potential constants were determined by fitting the frequency distribution of the $Zn^{2+}\cdots X$ distance (where X represents each of the 5 atoms coordinated to the Zn^{2+}) of a short unrestrained molecular dynamics run of MST1 with Ser-AMP bound.

3.2.3 Umbrella Sampling Calculations.

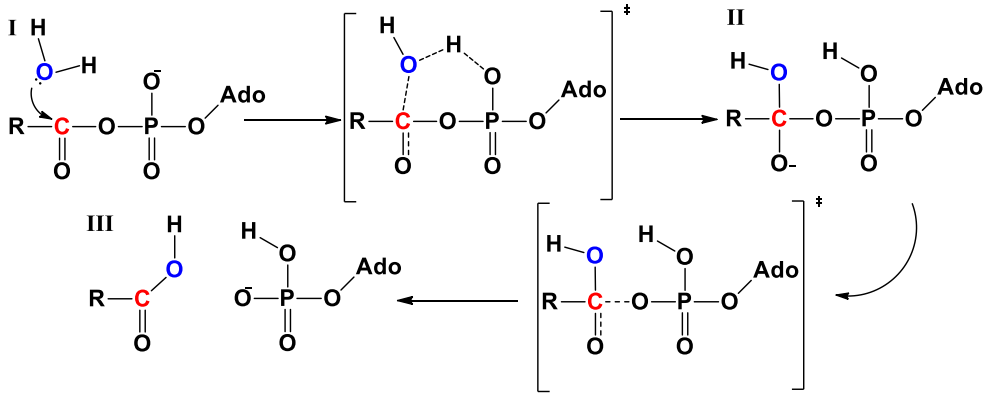
Preparation of the System. For the complexes of MST1 with bound aa-AMP three frames were extracted from each molecular dynamics trajectory according to criteria discussed below.

Construction of the Reaction Coordinates. The hydrolysis of aa-AMP in MST1 using umbrella sampling was investigated with a two-dimensional potential of mean force (PMF) followed by a one-dimensional PMF. In the two-dimensional reaction, the nucleophilic attack of the water oxygen against the carbonyl centre of the aa-AMP was described by the first reaction coordinate (RC1) corresponding to shortening of the d_2 distance in Scheme 3.2. The second reaction coordinate (RC2) was defined to be the antisymmetric combination of distances d_3 and d_4; that is, proton transfer from the mechanistic water to the phosphate. When RC1 and RC2 were combined in a reaction mechanism, the intermediate that was formed corresponded to the formation of a tetrahedral carbon intermediate (Scheme 3.3). Consequently, a one-dimensional reaction coordinate, RC3, was needed to separate the amino acid from the AMP, which was described solely by the lengthening of d_1 (Scheme 3.2).

Scheme 3.2. Definition of the reaction coordinates used for umbrella sampling of the hydrolysis of the aa-AMP within MST1.



Scheme 3.3. The mechanism of pre-transfer editing examined in this study showing the (I) reactant; (II) tetrahedral carbon intermediate; and (III) product complexes with their respective interconnecting transition structures.



3.2.4 QM/MM-MD Calculations.

A 100 ps equilibration was applied for the QM/MM-MD prior to umbrella sampling. The reactive QM region was chosen to include residues which exhibited significant changes in Mulliken charges through the reaction coordinates as observed for a small QM-cluster model. Following the widely used convention in the literature, each bond between the side groups and α-carbons of each residue was defined as one QM/MM boundary.[43] The restraints imposed on the Zn^{2+} penta-coordination during the MD simulations were maintained. The QM region included the aa-AMP substrate, Zn^{2+}, and the side chains of Cys133, Asp182, His184, His319, and a mechanistic water molecule (Figure 3.1), and thus was composed of 69 and 72 atoms when Ser- or Thr-AMP bound, respectively. The remainder of the enzyme was placed in the MM region.QM/MM-MD simulations were performed using the AMBER14[34] program with the SANDER module. The MM region of MST1 was treated with the ff14SB force field parameters[38] while the QM region was modelled using the AM1/d-PhoT semi-empirical method.[44] AM1/d-PhoT, a derivation of the AM1 semiempirical method with specific parameter sets for H, O and

P atoms, was chosen based on a benchmark study of several QM methods, including recent semi-empirical Hamiltonians (Supporting Information).

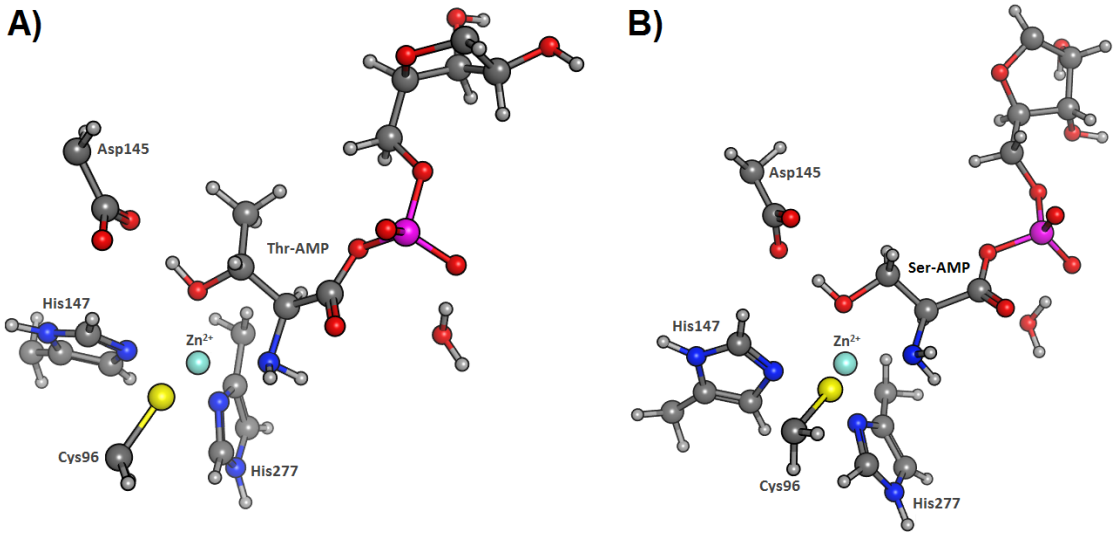


Figure 3.1. The QM region models used herein for (A) threonyl-adenylate and (B) seryl-adenylate bound within the MST1 active site.

Each umbrella sampling window was defined with a width of 0.10 Å, which resulted in 494 windows within a range of [RC1:RC2] = [3.2, 1.4: -1.5, 1]. In order to control the reaction coordinates, a harmonic potential centered on each window was applied with a force constant of 300 kcal mol^{-1} $Å^{-2}$. Each window was simulated during 15 ps with a time step of 0.5 fs. The results were analyzed using WHAM[45] to obtain the PMF of the pre-transfer editing mechanism.

3.3 Results and Discussion

3.3.1 Stability of the Simulation.

The root-mean-square deviations (RMSDs) of the position of the C_α atoms of the secondary structures of the catalytic domain of MST1 unbound (i.e., no ligand) and with Thr-AMP or Ser-AMP bound within its active site were monitored over the course of the

simulations with respect to the reference crystal structure, and are plotted in Figure 3.2. For all three systems the RMSD values during the simulations stayed within 2 Å of the initial reference structure, and were stable.

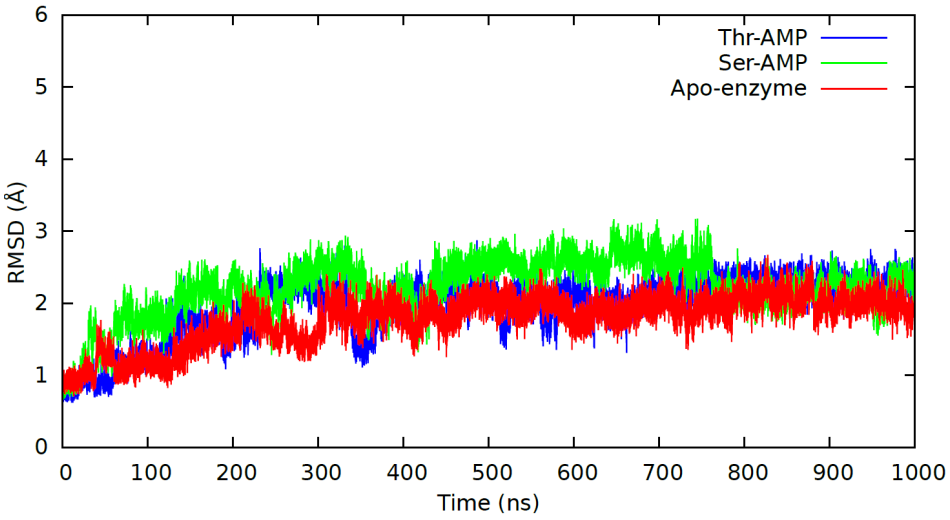


Figure 3.2. Plot of RMSDs of unbound MST1 (red) and when Thr-AMP (blue) or Ser-AMP (green) bound within its active site (see text).

3.3.2 Conformational Analysis of bound and solvated aa-AMP.

The root-mean square deviations (RMSD) of Thr-AMP and Ser-AMP position in bulk solution and when bound within the catalytic domain of MST1 were calculated with respect to the C_α atoms in the MST1 catalytic domain, excluding unordered secondary structure, were taken for comparison to the reference crystal structure (Figure 3.3). The RMSD of both Thr-AMP and Ser-AMP in bulk solution deviated significantly and fluctuated widely between 1.0-3.5 Å. In contrast, when bound, they show significantly smaller deviations and fluctuations in RMSD values of between 0.1-1.0 Å and 0.2-2.0 Å for Thr-AMP and Ser-AMP, respectively. This means that the flexibility of the aa-AMPs is restricted markedly when bound within the active site compared to when in bulk solution. More importantly, however, MST1 constrains the aa-AMP to a conformation

that is not observed in solution. Interestingly, Ser-AMP fluctuated briefly at an RMSD of approximately 1.7 Å which correlated to a 180° rotation of the N-C_{α}-C=O dihedral angle. In contrast, such a rotation was not observed for Thr-AMP. These results suggest that while MST1 constrains the positioning of the bound substrates, and in a conformation that is not reached in solution, non-cognate Ser-AMP is more mobile than cognate Thr-AMP within its active site due to its slightly smaller side chain.

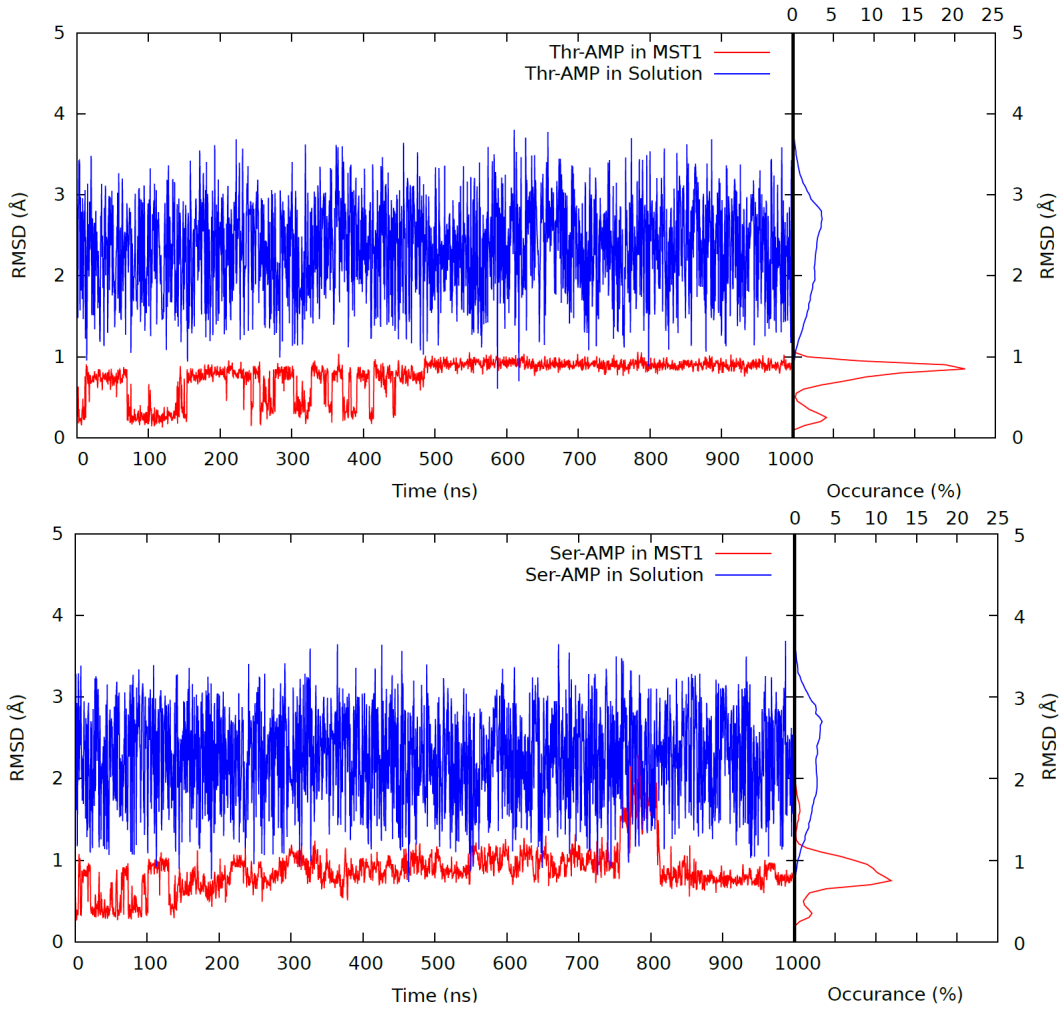


Figure 3.3. Plots of the RMSDs of threonyl-adenylate (top) and seryl-adenylate (bottom) position in bulk solution (blue) and when bound in the active site of MST1 (red), with respect to the reference heavy atoms in the crystal structure. The chart on the right side indicates the frequency of each RMSD conformation.

3.3.3 Active site···Thr-/Ser-AMP Interactions.

To see whether the differences in hydrolysis of MST1-bound Thr- and Ser-AMP were caused by conformational disparity, we monitored important selected hydrogen bonding interactions throughout each simulation. A ligand interaction map was generated, for each MD simulation, with percentage occurrence, and is shown in Figure 3.4. Since Thr-AMP and Ser-AMP differ chemically by only a -CH_2-, similarities were expected. As well, the interactions of the Zn^{2+} with the thiolate sulfur of Cys133, imidazole Nε and Nδ nitrogens of His184 and His319 respectively, and the amino acid substrates **α**-amino nitrogen and side chain hydroxyl oxygen were also maintained in both cases.

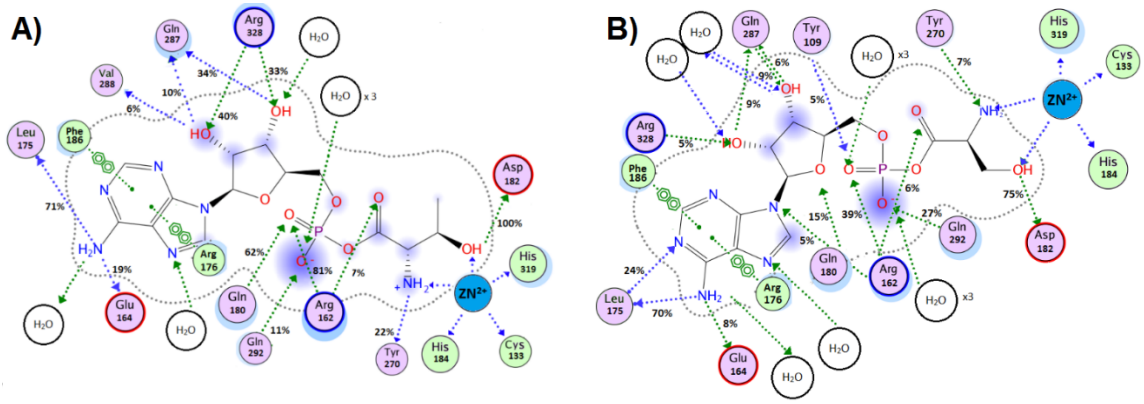


Figure 3.4. Ligand Interaction maps, with percentage occurrence of each interaction over the course of the MD simulation, for active site bound Thr-AMP (left) and Ser-AMP (right).

Active site···Adenine Interactions. The adenine nucleobases of both bound aa-AMP ligands lie sandwiched between Phe186 and Arg176 with which they interact through **π**-**π** stacking. In addition, both the –N6H_2 group quite consistently hydrogen bonds (~70% occurrence) with the backbone oxygen of Leu175 and, though much less consistently, with Glu164 and a water (see below). A water is also seen to hydrogen bond (2) quite frequently with the adenine N7 center. Several significant differences in active site-

substrate interactions are observed for bound Thr- and Ser-AMP. In particular, the backbone NH proton of Leu175 forms a H-bond with the N1 center of Thr-AMP at a much lower frequency (<5%) than with Ser-AMP (24%). Additionally, the Glu164 side chain carboxylate H-bonds with the $–N6H_2$ group of Thr-AMP more often (19%) than in the case of Ser-AMP (8%). Meanwhile, the guanidinium of Arg162 was observed to interact, though infrequently, with the adenine's N9 center in the case of Ser-AMP (5%), but not in the case of Thr-AMP (<5%).

Active site···Ribose Interactions. The guanidine group of Arg328 interacted with the 2'- (40%) and 3'-hydroxyl oxygens (33%) of Thr-AMP significantly more than its interaction with the 2'- (5%) and 3'-OH oxygens (<5%) of Ser-AMP. The ε oxygen of Gln287 interacted similarly with 2'-OH of Thr-AMP (10%) compared to Ser-AMP (9%). However, its interaction with 3'-OH of Thr-AMP (34%) was much greater than Ser-AMP (6%). The backbone oxygen of Val288 was seen to interact with 2'-OH of Thr-AMP (6%), but not with Ser-AMP (<5%). Meanwhile, the NεH proton of Gln180 stabilized the ribose ring oxygen of Ser-AMP (15%), but not that of Thr-AMP. Interestingly, waters were found to interact by hydrogen bonds to both the 2'- and 3'-OH groups of the Ser-AMP ribose, but with only one of these in the case of Thr-AMP.

Active site···Phosphate Interactions. The Arg162 guanidinium is observed to frequently H-bond with a phosphate oxygen of Thr-AMP (81%) but markedly less-so in the case of Ser-AMP (39%). Based on its H-bonds with the active site glutamyl residues Gln180 and Gln292, the Thr-AMP appears to be bound in two slightly different conformations (Figure S1). More specifically, the AMP phosphate frequently switches between H-bonding with the side chain $-NH_2$ groups of Gln180 and Gln292. In fact, the average RMSD values of 0.1 and 0.9 correspond to the H-bond interaction of the phosphate group with Gln180 and Gln292, respectively (Figure S1). However, the

longevity of the $_{\text{Thr-AMP}}PO\cdots H_2N_{Gln180}$ H-bond interaction throughout the simulation seems to indicate that this conformation is more stable.

Ser-AMP, on the other hand, was observed to possess a greater flexibility and wider range of conformations due in part to differences in its interactions with active site residues (Figure 3.4). In particular, although the phosphate of Ser-AMP did at times H-bond with the side chain -NH_2 of Gln292, it rarely H-bonded with that of Gln180 (<5%). Furthermore, Gln180 and Gln292 did not alternate in their H-bonding to the Ser-AMP phosphate (Figure S1). In fact, both Gln residues were frequently seen to dissociate their H-bonding from the phosphate group simultaneously. Consequently, water molecules were able to enter and replace these H-bonding networks and as a result, MST1-bound Ser-AMP had increased variability in its positioning within the active site compared to Thr-AMP.

Active site···Aminoacyl (Thr- and Ser-) Interactions. In addition to the above mentioned ligations of the aminoacyl (aa) moieties of the substrates to the Zn^{2+} ion, several other enzyme···aa interactions were observed. Similarities between Thr-AMP and Ser-AMP were observed for H-bonding between their carbonyl oxygen and the guanidinium of Arg162 which, in both cases was infrequent; 7% and 6%, respectively. In contrast, a considerably more consistent H-bond interaction was observed for both Thr- and Ser-AMP between their side chain hydroxyl group and the side chain carboxylate O_{carb} of Asp182. For Thr-AMP this interaction was observed throughout the simulation (i.e., 100%) while for Ser-AMP it was markedly less consistent at 75%. It is also noted that a H-bond was intermittently formed between the **α**-amino group of the Thr- and Ser- and the side chain hydroxyl of Tyr270 though with markedly greater occurrence for Thr-AMP (22%) compared to Ser-AMP (7%).

In general, the enzyme-substrate H-bonding interactions observed were longer lived (great consistency) for Thr-AMP than for Ser-AMP. For the latter, waters were observed to have greater accessibility in the active site and replaced these H-bonds (Figure 3.4).

3.3.4 Water Density Analysis.

Understanding the density and distribution of waters around each aa-AMP gives important information about its accessibility and potential availability for the required hydrolytic editing mechanism. Grid Inhomogeneous Solvation Theory (GIST)[46] was used to determine the density of water molecules, based on the waters oxygens, around both Thr-AMP and Ser-AMP. The iso-values, defined as the ratio of the local water density to the bulk water density,[46] associated with densities larger than 1 and 3 were mapped to identify any detectable differences in water accessibility within the active site (Figure 3.5). As can be seen, the Thr- and Ser-AMP bound active sites were both observed to be water accessible. Furthermore, in both cases water-rich regions were found adjacent to the substrates carbonyl group, phosphate oxygens, and their **α**-amine groups.

However, several key differences were observed between the bound complexes. In particular, in the case of Thr-AMP water was only accessible on the solvent side while for Ser-AMP water was accessible on both sides. This is likely due to the smaller side chain of Ser-AMP and its greater mobility leading to the formation of a void which allows water molecules to fill the gap (Figure 3.5b and 3.5d). Due to this accessibility of water on the distal side of Ser-AMP away from the solvent, and unlike that observed for Thr-AMP (Figure 3.5A and 5B), there were regions of high water density on both sides of its backbone carbonyl group (i.e., C_{carb}=O), a mechanistically key region. As a result, and unlike for Thr-AMP, it is possible that nucleophilic attack of a water on the ligand, i.e., hydrolysis, may be initiated from either side of the ligands C_{carb} center. Previously, Ling

et al.[28] had noted that in their X-ray crystallographic structures waters were positioned slightly differently for when Thr-AMP and Ser-AMP were bound. In this present study differences were indeed observed in the positioning and availability of active site waters for the two aa-AMP ligands. Additionally, the increased accessibility also enabled the water molecules to participate in a greater number of hydrogen bonds with the ligand, in particular its carbonyl group.

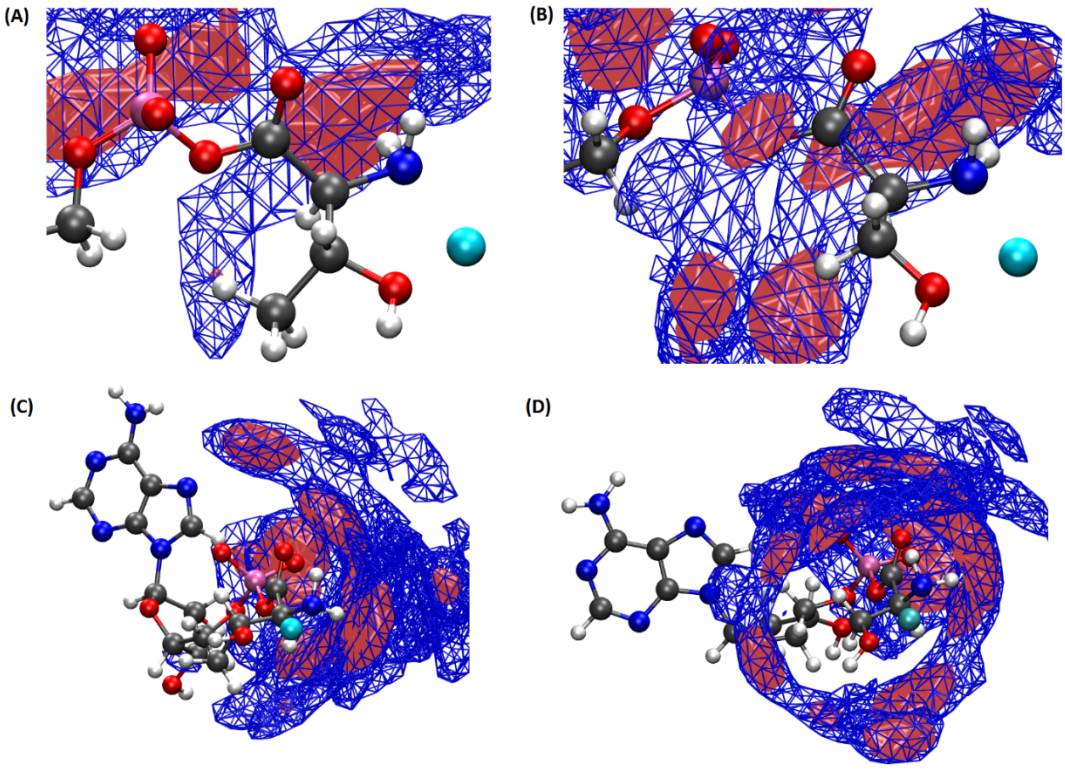


Figure 3.5. The water density surrounding bound Thr-AMP (A and C) and Ser-AMP (B and D). Blue and red denotes regions of medium (iso-value = 1) and high (iso-value = 3) water density, repsectively. Analyses were done using GIST.[46]

3.3.5 Kinetic Aspects: Umbrella Sampling.

In order to elucidate the mechanism of Thr-AMP and Ser-AMP hydrolysis, we performed umbrella sampling on both ligands from 3 different initial conformations. They

were chosen based on the RMSD of the three most populated conformers. These structures were taken from t = 200, 440, and 850 ns for Thr-AMP (hereafter referred to as T1, T2, and T3) and from t = 100, 780, 900 ns for Ser-AMP (hereafter referred to as S1, S2, S3). These structures were subjected to a 2D umbrella sampling prior to a 1D umbrella sampling.

Nucleophilic attack of the Water. The 2D free energy surfaces obtained using QM/MM-MD umbrella sampling for when Thr-AMP and Ser-AMP are bound in the MST1 active site are shown in Figure 3.6. Both Thr-AMP and Ser-AMP were energetically amenable to hydrolysis. Interestingly, T1, T2, and T3 were similar to each other thermodynamically and kinetically, while S1, S2, and S3 were also very similar to each other. Hence, positional differences of these ligands do not appear to greatly influence its hydrolysis and as a result, each of these ligand conformations can be considered as near attack conformers (NAC).

In the first step of hydrolysis, the minimum energy path corresponds to an initial decrease in RC1, followed by an increase in RC2. Chemically, this indicates that the oxygen of the mechanistic water initiates a nucleophilic attack on the carbonyl carbon of the aa-AMP substrate (RC1). Subsequently, and once the attack is in progress, the water transfers a proton onto a nearby oxygen of the phosphate group (RC2). This results in formation of a tetrahedral carbon intermediate, an oxyanion, with a neutral phosphate group. The reactive complex (RC) was located for all systems at approximately [RC1:RC2 = (3.2, -1.1)] while formation of the intermediate complex (IC) occurred at [RC1:RC2 = (1.5:1.0)]. A first order-saddle point (transition structure (TS)) for Thr-AMP was detected around [RC1:RC2 = (1.5:-0.2)]. In contrast, in the case of Ser-AMP, two TS's were located: the first, TS', at [RC1:RC2 = (1.8,-1.1)] and the second, TS", at about [RC1:RC2 = (1.5:-0.2)]. Between these two transition structures a metastable intermediate was located at around [RC1:RC2 = (1.6,-1.0)]. In terms of reaction

coordinates, TS' is associated with the nucleophilic attack of the water on C_{carb} center of aa-AMP while TS" is associated with the proton transfer from the attacking water onto a non-bridging oxygen of the substrates AMP phosphate. For Thr-AMP both of these processes occur essentially concomitantly (see below). Indeed, it should be noted that the location along the reaction coordinate of TS in the first step of hydrolysis of Thr-AMP, is the approximate same position as TS" in the mechanism for Ser-AMP (see also Figure 3.7).

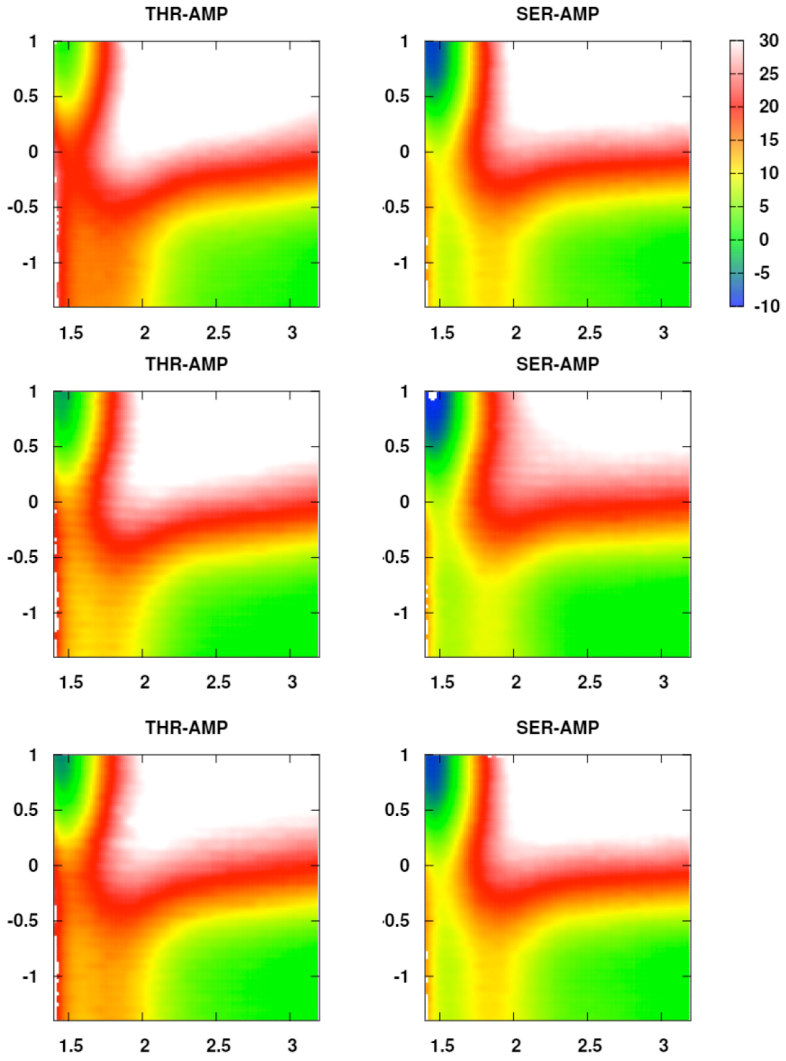


Figure 3.6. Plots of the free energy surfaces (kcal mol^{-1}) obtained from 2D umbrella sampling for the first step in the hydrolysis of the three MST1-bound conformers of Thr-AMP (left; T1-T3 (top to bottom)) and Ser-AMP (right; S1-S3 (top to bottom)). Free

energies are reported relative to the corresponding reactant complex (RC) located at [RC1, RC2] = [3.2, -1.5].

From Figure 3.6, it can be seen that at least when bound within the active site the first step of hydrolysis, formation of a tetrahedral intermediate, is kinetically and thermodynamically favoured for Ser-AMP compared to Thr-AMP . To quantify free energy values, a 1D reaction coordinate surface that goes through RC and all intermediates and transition states, up to formation of the tetrahedral intermediate was derived and shown in Figure 3.7. This illustrates the free energy variation obtained along the reaction coordinate of the first step of hydrolysis for all six conformers (T1-3 and S1-3) of bound MST1-substrate complexes considered. The initial attack of the water on Thr-AMP and Ser-AMP occurs with a barrier of 14-18 kcal mol^{-1} and 8-10 kcal mol^{-1} respectively, and is in fact the overall mechanistic rate limiting step (RLS) for both. Although, as noted below, for Ser-AMP the second step of the overall mechanism has a similar free energy barrier. It should be noted that for all conformations considered, the barrier for this step was higher for Thr-AMP than Ser-AMP by approximately 4-11 kcal mol^{-1}.

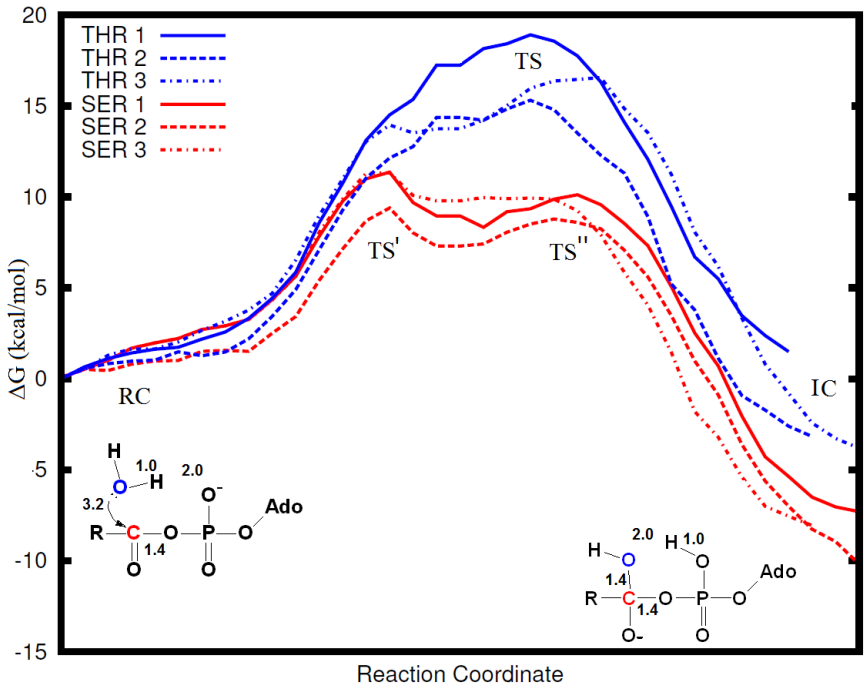


Figure 3.7. Thermodynamic properties of the first step of the pre-transfer editing mechanism comparing Thr-AMP to Ser-AMP. All free energy differences are relative to the reaction complexes (RC). Values were extracted by a 1D projection of the 2D umbrella sampling results along the minimal free energy pathway (see Figure 3.6). Distances are denoted in Å.

Extraction of frames from umbrella sampling windows around [RC1,RC2 = (1.6,-1.0)] show an important difference between Thr-AMP and Ser-AMP reaction pathways. In the latter, the metastable intermediate was stabilized by H-bonds between its oxyanion center and two waters. In contrast, for Thr-AMP there were at most only one water (if any) present to stabilize the charge build-up on the oxygen (Figure 3.8). A similar trend is observed for the corresponding tetrahedral intermediate complexes (IC) shown in Figure 3.9. For the IC derived from Ser-AMP, three waters can be seen H-bonded to the oxyanion center. In contrast, for the IC derived from Thr-AMP there are at most one to two water molecules H-bonded to the oxyanion center. This greater stabilization of IC for hydrolysis of Ser-AMP compared to Thr-AMP is also observed energetically: in

hydrolysis of Thr-AMP the IC is approximately isoenergetic with the corresponding RC whereas for Ser-AMP IC lies lower in energy than its corresponding RC by 7-10 kcal mol^{-1}. The increased water accessibility for the case of bound Ser-AMP enabled waters to participate in a greater number of hydrogen bonds with the ligand, in particular its carbonyl group. As a result, hydrolytic editing of Ser-AMP is kinetically and thermodynamically favoured over that of Thr-AMP due to greater stabilization of the negative charge build-up on the substrates' carbonyl oxygen, and corresponding oxyanion intermediate during the mechanism.

Thus the occurrence of a metastable intermediate in nucleophilic attack of a water on Ser-AMP appears to be due to the presence of waters on the opposite side of ligand from which the hydrolytic water is attacking, that help stabilize the forming oxyanion. This greater solvent driven stabilization also results in the lower free energy barrier described above for nucleophilic attack of the water on Ser-AMP than that obtained for Thr-AMP. It is noted that for both ligands the reverse reaction has similar free energy barriers.

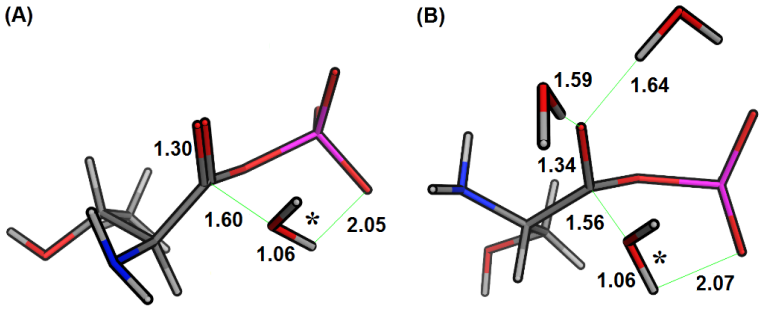


Figure 3.8. Representative structures extracted from [RC1, RC2] = [1.6, -1.0] for MST1-bound (A) Thr-AMP and (B) Ser-AMP. The hydrolytic water is denoted by an asterisk (*).

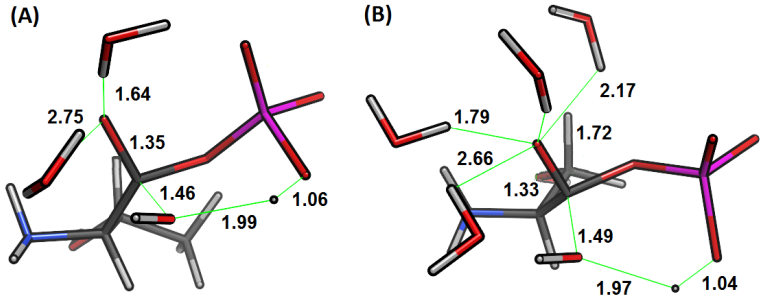


Figure 3.9. Representative structures of the tetrahedral intermediate complex, extracted from [RC1, RC2] = [1.5, 1.0], for hydrolysis of MST1-bound (A) Thr-AMP and (B) Ser-AMP.

Cleavage of the C_{carb}-OP bond and product formation. The distance between the carbonyl carbon (C_{carb}) of the amino acid moiety of the aa-AMP substrate and the bridging oxygen of the leaving phosphate group (Scheme 2; d_1) was elongated from 1.4 Å to 2.5 Å (Figure 3.10). Free energy minima corresponding to the intermediate complex (IC) and product complex (PC) were located at d_1 = 1.5 Å and 2.4 Å, respectively. For both hydrolytic mechanisms (Thr- and Ser-AMP) the transition structures for this cleavage of the C_{α}-OP bond ($TS^{(2)}$) were located at d_1 = 1.95 Å and were approximately 7-12 kcal mol^{-1} higher for Thr-AMP and 1-5 kcal mol^{-1} higher for Ser-AMP in free energy than their corresponding RC (Figure 3.10). Interestingly, no significant structural differences were seen for the separation of the amino acid from AMP.

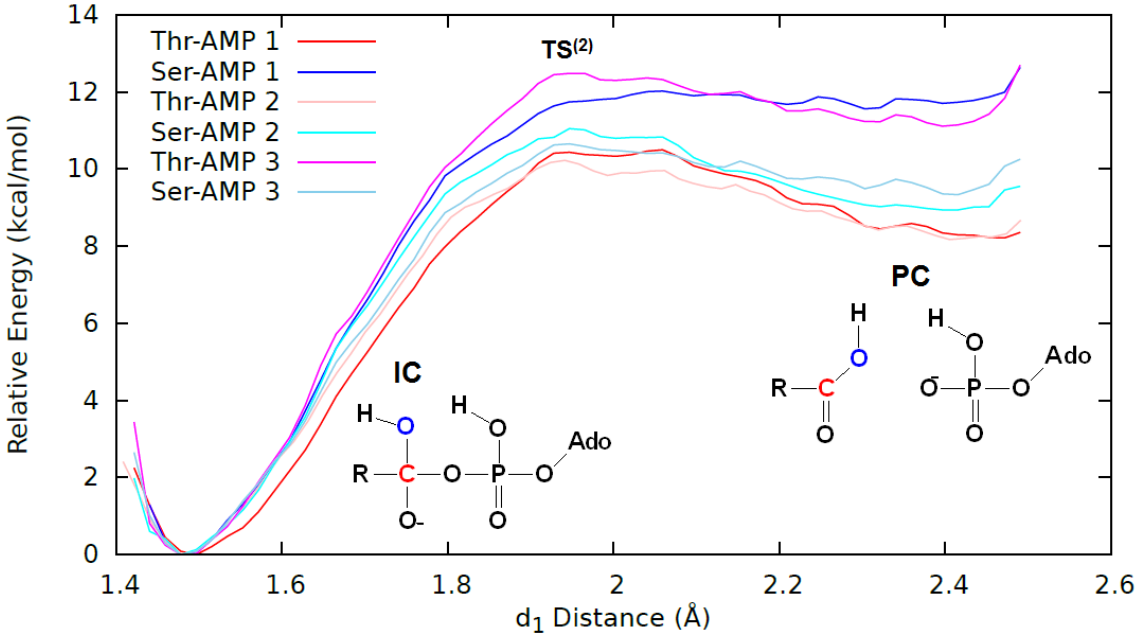


Figure 3.10. A 1D umbrella sampling of the elongation of the $_{AMP}PO–C_{Carb}$ distance (d_1).

The successful formation of the product complex was validated by performing an improper angle measurement (CA-OH-O-C improper angle: Figure S7) for the entire duration of the sampling window corresponding to d_1 = 2.4 Å (Figure S8). The results obtained indicate that this improper angle fluctuates by ±5° for all samplings (Figure S9). The resulting product complexes (PC) obtained for Thr-AMP and Ser-AMP after umbrella sampling were higher in relative free energy than their corresponding reactant complexes (RC) by 10 and 5 kcal mol^{-1}, respectively. Thus, the products derived from hydrolysis of Ser-AMP are thermodynamically more favoured compared to those of Thr-AMP. It is noted that in the present study the PCs are in their carboxylic acid forms and thus, further stabilization may be possible via the loss of the acidic proton to solution.

Previously, Ling et al.[28] had proposed that, based on the crystallographic structures they obtained that Asp112 may act as a base to activate Tyr109, which in turn deprotonates the water that nucleophilically attacks the substrates phosphorus (P) center during hydrolytic editing. However, this seems unlikely given that the distance between side chain oxygen of Tyr109 and P and C_{carb} centers of Thr- and Ser-AMP are both consistently over 5 Å during the course of the MD simulation. Thus, Tyr109 is not correctly positioned to facilitate nucleophilic attack of the water at either of these centers. Additionally, the deprotonation of water, as our results show, should occur after its nucleophilic attack of the C_{carb} centre. It is noted that the computational study of the post-translational editing mechanism of leucyl-tRNA synthetase by Matsuno et al.[47] similarly observed that deprotonation of the water occurred after it had begun nucleophilic attack on the C_{carb} center of the misacylated aminoacyl-tRNA (aa-tRNAs).

3.3.6 Role of MST1 in Pre-transfer Editing.

Previous experimental work[28] had suggested that pre-transfer editing could occur via two main possible pathways. In one pre-transfer editing occurs in within the

aminoacylation site of MST1 where Ser-AMP is preferentially hydrolysed at a rate 4-times compared to Thr-AMP. In the second potential pathway the aa-AMP ligands are released into the aqueous environment whereby they both could be hydrolyzed. However, this latter process occurs an order of magnitude slower than in the aminoacylation active site and fails to sufficiently discriminate between the two ligands.[28] The present computational study appears to support the above experimental suggestion that the aminoacylation active site of MST1 is crucial and required for discriminating the non-cognate substrate Ser-AMP from the cognate substrate Thr-AMP both through greater conformational restriction of Thr-AMP and increased water accessibility for bound Ser-AMP. Furthermore, pre-transfer editing proceeds via a substrate-assisted mechanism in which the AMP phosphate acts as a base to help activate the water nucleophile.

3.4 Conclusion

In this study we have computationally investigated the binding of the cognate and non-cognate substrates Thr-and Ser-AMP respectively within the aminoacylation active site of yeast mitochondrial threonyl-tRNA synthetase (MST1), and their subsequent pre-transfer editing mechanisms. More specifically, microsecond MD simulations have been used to examine the unbound and bound MST1···aa-AMP complexes, while umbrella sampling and QM/MM-MD have been applied to elucidate the mechanism of pre-transfer editing. In order to elucidate the preference of MST1 for editing against one substrate over the other, particular attention was paid to three descriptors: (i) conformational analysis of the unbound and bound ligands; (ii) key enzyme-substrate hydrogen bonding interactions; and (iii) the water density around the bound ligands.

Importantly, it was observed that when bound within the MST1 aminoacylation active site both aa-AMPs were constrained in a conformation distinct from that observed when they were in the bulk solution. Furthermore, this positioning appears to render both

ligands more susceptible to hydrolysis than in bulk solution. Despite Thr-AMP and Ser-AMP differing by just an additional methylene (the extra -CH2- being in the side chain of the threonyl aminoacyl moiety), the MST1-bound Thr-AMP appears to be more conformationally rigid than Ser-AMP. Water density analysis of the MST1-bound complexes indicated that, as a result of this greater variability and smaller side chain, water is better able to permeate the aminoacylation active site when Ser-AMP is bound. Indeed, water was observed to surround the key phosphate-carbonyl region of Ser-AMP. In contrast, when Thr-AMP is bound water was observed to only permeate on one side of the substrate. This greater availability of water in the case of bound Ser-AMP not only increased the chances of water being present for hydrolysis to occur, but also provided greater stabilization of the oxyanion center and tetrahedral intermediate formed during the course of the reaction (see below).

The hydrolytic pre-transfer editing of both Thr-AMP and Ser-AMP occurred via a two-step process. More specifically, it is initiated by nucleophilic attack of a water molecule at the carbonyl carbon (C_{carb}) center of the aa-AMP substrate, with the phosphate group acting as the base, to form an oxyanion tetrahedral intermediate. That is, pre-transfer editing proceeds via a substrate-assisted mechanism. Importantly, this first step yielded the only major mechanistic differences between the two ligands. In particular, the free energy barrier for formation of the tetrahedral intermediate was significantly greater for Thr-AMP than Ser-AMP by 4-10 kcal mol^{-1}, which is likely sufficient to kinetically discriminate between these two ligands. Furthermore, for Thr-AMP this process was the clear overall rate limiting step for its hydrolytic pre-transfer editing by MST1.

The structures of the TS and tetrahedral intermediate regions were examined for both ligands. In particular, it was observed that in the case of Ser-AMP the oxyanion center and tetrahedral intermediate are stabilized by at least two or more hydrogen bonds with

active site waters. In contrast, for Thr-AMP only at most one water was observed to form an analogous stabilizing hydrogen bond with the oxyanion or intermediate.

The subsequent cleavage of the $_{AMP}PO–C_{Carb}$ bond to give the amino acid and AMP was determined to occur for Ser-AMP and Thr-AMP with quite similar free energy barriers to each other. In the case of Ser-AMP this barrier was also in reasonable agreement with the free energy barrier for the first reaction step, nucleophilic attack of a water and formation of a tetrahedral intermediate. Thus, for Thr-AMP this second step in the mechanism has a much lower barrier (by ~7-10 kcal mol^{-1}) than the first. The subsequent product complexes (PC), MST1 with the amino acid and AMP bound in its active site, resulting from hydrolysis of Ser-AMP was thermodynamically much lower in energy by approximately 5-10 kcal mol^{-1} than those resulting from Thr-AMP. This greater thermodynamic stability of the Ser-AMP derived product complexes is possibly due at least in part to the greater number of hydrogen bonds the now cleaved serine component makes with water (three) compared to the cleaved threonine (two).

The markedly higher overall barrier for hydrolytic editing of the cognate Thr-AMP substrate in the aminoacylation active site compared to that of the non-cognate substrate Ser-AMP is likely sufficient for enabling MST1 to accurately discriminate between them.

3.5 Associated Content

Supporting Information.

A benchmark study of semiempirical methods for studying the MST1 pre-transfer editing mechanism. The supporting information is available in Appendix A

3.6 References

(1) Martinis, S. A.; Boniecki, M. T. The balance between pre- and post-transfer editing in tRNA synthetases. *FEBS Lett.* **2010**, *584*, 455-459.

(2) Bushnell, E. A. C.; Huang, W. J.; Llano, J.; Gauld, J. W. Molecular Dynamics Investigation into Substrate Binding and Identity of the Catalytic Base in the Mechanism of Threonyl-tRNA Synthetase. *J. Phys. Chem. B* **2012**, *116*, 5205-5212.

(3) Huang, W. J.; Gherib, R.; Gauld, J. W. An Active Site Water Broadens Substrate Specificity in S-Ribosylhomocysteinase (LuxS): A Docking, MD, and QM/MM Study. *J. Phys. Chem. B* **2012**, *116*, 8916-8929.

(4) Sankaranarayanan, R.; Dock-Bregeon, A. C.; Romby, P.; Caillet, J.; Springer, M.; Rees, B.; Ehresmann, C.; Ehresmann, B.; Moras, D. The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* **1999**, *97*, 371-381.

(5) Halwani, R.; Cen, S.; Javanbakht, H.; Saadatmand, J.; Kim, S.; Shiba, K.; Kleiman, L. Cellular distribution of Lysyl-tRNA synthetase and its interaction with gag during human immunodeficiency virus type 1 assembly. *J. Virol.* **2004**, *78*, 7553-7564.

(6) Stark, L. A.; Hay, R. T. Human immunodeficiency virus type 1 (HIV-1) viral protein R (Vpr) interacts with Lys-tRNA synthetase: Implications for priming of HIV-1 reverse transcription. *J. Virol.* **1998**, *72*, 3037-3044.

(7) Wellman, T. L.; Eckenstein, M.; Wong, C.; Rincon, M.; Ashikaga, T.; Mount, S. L.; Francklyn, C. S.; Lounsbury, K. M. Threonyl-tRNA synthetase overexpression correlates with angiogenic markers and progression of human ovarian cancer. *BMC Cancer* **2014**, *14*.

(8) Randau, L.; Schauer, S.; Ambrogelly, A.; Salazar, J. C.; Moser, J.; Sekine, S.; Yokoyama, S.; Soll, D.; Jahn, D. tRNA recognition by glutamyl-tRNA reductase. *J. Biol. Chem.* **2004**, *279*, 34931-34937.

(9) Howard, O. M.; Dong, H. F.; Yang, D.; Raben, N.; Nagaraju, K.; Rosen, A.; Casciola-Rosen, L.; Hartlein, M.; Kron, M.; Yang, D.; Yiadom, K.; Dwivedi, S.; Plotz, P. H.; Oppenheim, J. J. Histidyl-tRNA synthetase and asparaginyl-tRNA synthetase,

autoantigens in myositis, activate chemokine receptors on T lymphocytes and immature dendritic cells. *J. Exp. Med.* **2002**, *196*, 781-791.

(10) Mathews, M. B.; Reichlin, M.; Hughes, G. R.; Bernstein, R. M. Anti-threonyl-tRNA synthetase, a second myositis-related autoantibody. *J. Exp. Med.* **1984**, *160*, 420-434.

(11) Cvetesic, N.; Perona, J. J.; Gruic-Sovulj, I. Kinetic Partitioning between Synthetic and Editing Pathways in Class I Aminoacyl-tRNA Synthetases Occurs at Both Pre-transfer and Post-transfer Hydrolytic Steps. *J. Biol. Chem.* **2012**, *287*, 25381-25394.

(12) Liu, H. N.; Gauld, J. W. Substrate-assisted Catalysis in the Aminoacyl Transfer Mechanism of Histidyl-tRNA Synthetase: A Density Functional Theory Study. *J. Phys. Chem. B* **2008**, *112*, 16874-16882.

(13) Wu, J.; Fan, Y. Q.; Ling, J. Q. Mechanism of oxidant- induced mistranslation by threonyl- tRNA synthetase. *Nucleic Acids Res.* **2014**, *42*, 6523-6531.

(14) Perona, J. J.; Gruic-Sovulj, I. Synthetic and editing mechanisms of aminoacyl-tRNA synthetases. *Top. Curr. Chem.* **2014**, *344*, 1-41.

(15) Zhou, X. L.; Ruan, Z. R.; Wang, M.; Fang, Z. P.; Wang, Y.; Chen, Y.; Liu, R. J.; Eriani, G.; Wang, E. D. A minimalist mitochondrial threonyl-tRNA synthetase exhibits tRNA-isoacceptor specificity during proofreading. *Nucleic Acids Res.* **2014**, *42*, 13873-13886.

(16) Hagiwara, Y.; Field, M. J.; Nureki, O.; Tateno, M. Editing Mechanism of Aminoacyl-tRNA Synthetases Operates by a Hybrid Ribozyme/Protein Catalyst. *J Am Chem Soc* **2010**, *132*, 2751-2758.

(17) Perona, J. J.; Gruic-Sovulj, I. Synthetic and Editing Mechanisms of Aminoacyl-tRNA Synthetases. *Aminoacyl-Trna Synthetases in Biology and Medicine* **2014**, *344*, 1-41.

(18) Tardif, K. D.; Liu, M. S.; Vitseva, O.; Hou, Y. M.; Horowitz, J. Misacylation and editing by Escherichia coli Valyl-tRNA synthetase: Evidence for two tRNA binding sites. *Biochemistry* **2001**, *40*, 8118-8125.

(19) Hagiwara, Y.; Nureki, O.; Tateno, M. Structural modelling of the complex of leucyl-tRNA synthetase and mis-aminoacylated tRNA(Leu). *FEBS Lett.* **2009**, *583*, 825-830.

(20) Jakubowski, H. Molecular basis of homocysteine toxicity in humans. *Cell. Mol. Life Sci.* **2004**, *61*, 470-487.

(21) ISI Web of Science with Search Term: "Pre-transfer & tRNA" vs "Post-transfer & tRNA". (accessed Aug 8, 2016 2016).

(22) Ling, J.; Peterson, K. M.; Simonovic, I.; Cho, C.; Soll, D.; Simonovic, M. Yeast mitochondrial threonyl-tRNA synthetase recognizes tRNA isoacceptors by distinct mechanisms and promotes CUN codon reassignment. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3281-3286.

(23) Lee, J. W.; Beebe, K.; Nangle, L. A.; Jang, J.; Longo-Guess, C. M.; Cook, S. A.; Davisson, M. T.; Sundberg, J. P.; Schimmel, P.; Ackerman, S. L. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature* **2006**, *443*, 50-55.

(24) Fersht, A. R. Editing mechanisms in protein synthesis. Rejection of valine by the isoleucyl-tRNA synthetase. *Biochemistry* **1977**, *16*, 1025-1030.

(25) Tan, M.; Zhu, B.; Zhou, X. L.; He, R.; Chen, X.; Eriani, G.; Wang, E. D. tRNA-dependent Pre-transfer Editing by Prokaryotic Leucyl-tRNA Synthetase. *J. Biol. Chem.* **2010**, *285*, 3235-3244.

(26) Beuning, P. J.; Musier-Forsyth, K. Species-specific differences in amino acid editing by class II prolyl-tRNA synthetase. *J. Biol. Chem.* **2001**, *276*, 30779-30785.

(27) Splan, K. E.; Ignatov, M. E.; Musier-Forsyth, K. Transfer RNA modulates the editing mechanism used by class II prolyl-tRNA synthetase. *J. Biol. Chem.* **2008**, *283*, 7128-7134.

(28) Ling, J. Q.; Peterson, K. M.; Simonovic, I.; Soll, D.; Simonovic, M. The Mechanism of Pre-transfer Editing in Yeast Mitochondrial Threonyl-tRNA Synthetase. *J. Biol. Chem.* **2012**, *287*, 28518-28525.

(29) Minajigi, A.; Francklyn, C. S. Aminoacyl Transfer Rate Dictates Choice of Editing Pathway in Threonyl-tRNA Synthetase. *J. Biol. Chem.* **2010**, *285*, 23810-23817.

(30) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pK(a) values for protein-ligand complexes. *Proteins-Structure Function and Bioinformatics* **2008**, *73*, 765-783.

(31) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK(a) values. *Proteins-Structure Function and Bioinformatics* **2005**, *61*, 704-721.

(32) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525-537.

(33) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284-2295.

(34) D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman. AMBER 2015. *University of California, San Francisco* **2015**.

(35) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(36) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542-1555.

(37) Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878-3888.

(38) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

(39) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247-260.

(40) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(41) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.;

Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision D.01. *Gaussian, Inc., Wallingford CT* **2009**.

(42) Andersen, H. C. Molecular-Dynamics Simulations at Constant Pressure and-or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384-2393.

(43) Fernando Clemente, T. V., and Michael J. Frisch: *Getting the Most out of ONIOM: Guidelines and Pitfalls*; Wiley-VCH, 2010.

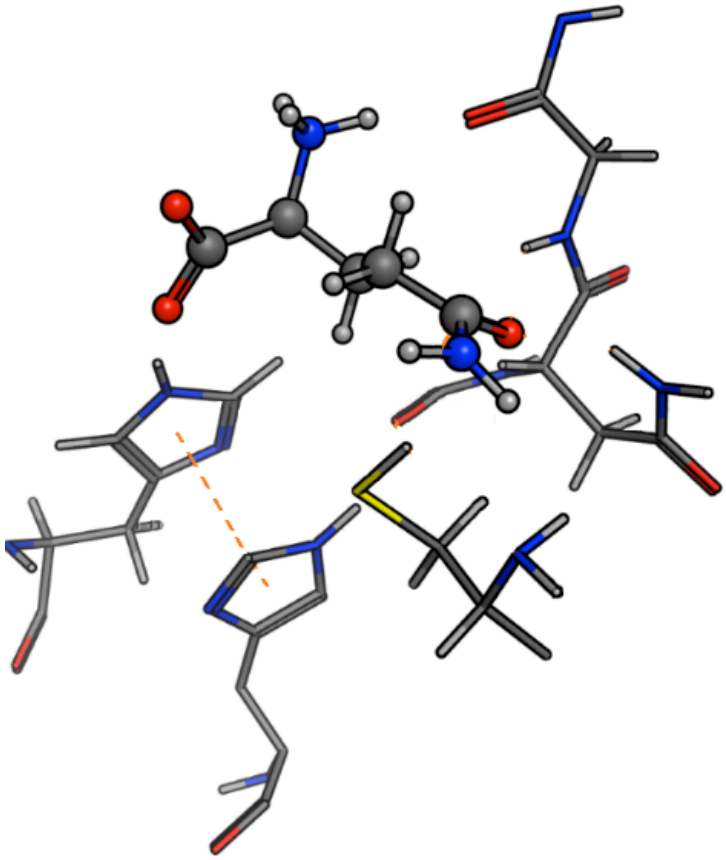
(44) Nam, K.; Cui, Q.; Gao, J. L.; York, D. M. Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms. *J. Chem. Theory Comput.* **2007**, *3*, 486-504.

(45) Grossfield, A. WHAM: The weighted histogram analysis method, version 2.0.9. *University of Rochester Medical Center, Rochester, NY*.

(46) Nguyen, C. N.; Young, T. K.; Gilson, M. K. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J. Chem. Phys.* **2012**, *137*.

(47) Hagiwara, Y.; Field, M. J.; Nureki, O.; Tateno, M. Editing mechanism of aminoacyl-tRNA synthetases operates by a hybrid ribozyme/protein catalyst. *J. Am. Chem. Soc.* **2010**, *132*, 2751-2758.

Chapter 4: A Critical Role for the Protonation State of the N-terminal Amine of Glucosamine-6-Phosphate Synthase (GlmS) on its Mechanism. A Computational Investigation



4.1 Introduction

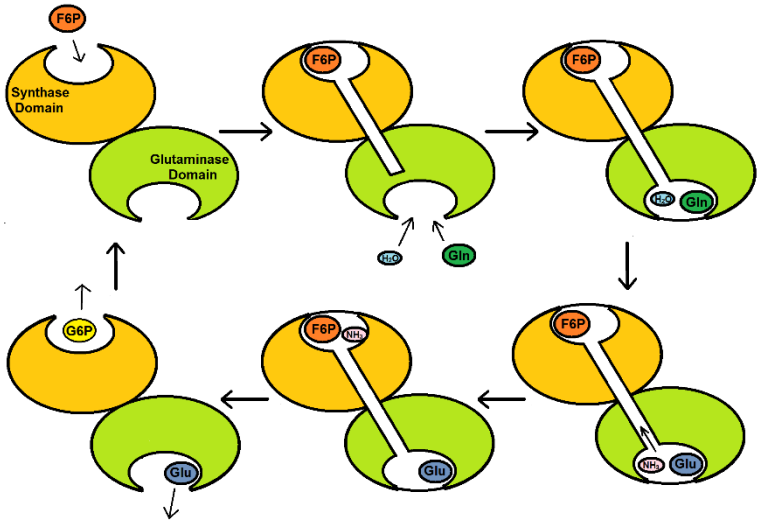
The class II glutamine-dependent amidotransferase (Gn-AT) family of enzymes is central to a variety of important physiological processes in a broad range of organisms from bacteria to mammals. They are multi-active site enzymes; each contains a functionally conserved glutaminase domain but possess a unique synthase domain. At least four known homologous proteins are included in this family of enzymes,[1] which are essential for the biosynthesis of purines,[2] asparagine,[3] glutamate,[4,5] and hexosamine.[6-8] A crucial Gn-AT enzyme involved in the latter pathway is glucosamine-6-phosphate synthase (GlmS). More specifically, it is responsible for synthesizing D-glucosamine-6-phosphate (GlcN-6-P) as one of its main products.[6]

GlcN-6-P is a key metabolic precursor for a plethora of important macromolecules in different organisms. For instance, in humans and other mammals it is an important precursor in the synthesis of a variety of glycoproteins, and as a sensory molecule of glucose uptake.[8] In fact, when adipocytes and fibroblasts were exposed to GlcN-6-P they developed insulin resistance[9] and expressed phenotypes similar to type II diabetes in transgenic mice.[7] Subsequent work on the inhibition of GlmS showed that insulin resistance could be prevented,[10] leading to a number of proposed enzyme inhibitors.[11] In bacteria, GlcN-6-P is the precursor of peptidoglycan and lipopolysaccharides (LPS) that are the building blocks of their cell walls. This has resulted in an interest in the development of different inhibitors against GlmS as potential antibiotics.[12-15] Similarly, in fungi, inhibitors of GlmS have gained interest because GlcN-6-P is the required substrate for chitin formation.[14] Experimentally, it has been shown that in fungi inhibition of this enzyme for even a short period of time is lethal. In contrast, short-term inhibition of human GlmS (Gfat) is not lethal because it is quickly re-expressed and its hexosamine products also have reasonably long lifespans.[15]

Unlike other Gn-ATs, GlmS cannot use exogenous ammonia as a source of nitrogen.[9] That is its overall synthetic role is critically dependent on the hydrolytic deamination of

glutamine in its glutaminase domain to produce free ammonia (Scheme 4.1).[8] The latter is then shuttled to the synthase domain, over 18 Å away, via a hydrophobic channel that helps prevent loss of the ammonia to the solvent.[16,17] In the synthase domain the NH_3 is reacted with D-fructose-6-phosphate (F6P) to produce GlcN-6-P. Protein crystallization with intermediate analogues have shown that GlmS catalysis is modular and occurs in a specific order (see Scheme 4.1).[16] F6P binding in the synthase domain triggers the glutaminase domain to bind L-glutamine and the formation of the ammonia channel, though it is blocked by the Q-loop secondary structure. However, the sealing of the glutaminase domain after L-glutamine binding allows the tunnel to connect. After the products are formed in both domains, the L-glutamic acid leaves first, followed by the GlcN-6-P. This tight regulation also ensures that the ammonia from the glutaminase domain is not lost to the medium.[16]

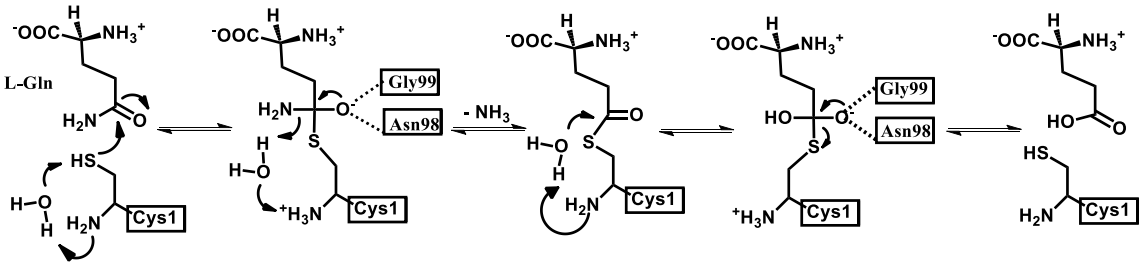
Scheme 4.1. The catalysis of different substrates (glutaminase and synthase domains) within GlmS occur in several distinct stages.



Site-directed mutagenesis and kinetics experiments have suggested that a conserved N-terminal cysteinyl (Cys1) thiolate is responsible for initiating the nucleophilic attack

against the amide carbon centre of the substrate.[16] Indeed, inhibition and mutational studies have shown that GlmS can be inactivated by covalent modification of the thiol of Cys1 by the glutamine analogue N3-(4-methoxyfumaroyl)-l-2,3-diaminopropanoic acid (FMDP).[14,16] In addition, previous pKa calculations using PROPKA[18] suggested that the buried N-terminal amine of Cys1 has a markedly reduced pKa (~5-5.3) and is likely neutral at physiological pH.[16] It has been further suggested based on X-ray crystallographic structures that the N-terminal Cys1 amine may be hydrogen bonded to the side chain hydroxyl of Thr606, thus stabilizing the neutral form.[6] Furthermore, mutational and kinetic studies have shown that glycylation of the N-terminus Cys1 amine of GlmS drastically reduces its glutaminase activity.[16] As a consequence, it has been proposed that the Cys1 residue could also act as a mechanistic base via its N-terminal amine group (Scheme 4.2).

Scheme 4.2. Proposed catalytic mechanism of the glutaminase domain of GlmS.[16]



More specifically, it has been proposed that the neutral **α**-amine of Cys1 helps activate the thiol of Cys1 via a water molecule that bridges the two groups. The resulting thiolate is then able to nucleophilically attack the glutamine substrate to form a covalently cross-linked tetrahedral oxyanion intermediate. It should be noted that Asn98 and Gly99 have been suggested to form an oxyanion hole to help stabilize the negative charge buildup on the substrate's oxygen during the reaction.[19] The subsequent collapse of the

tetrahedral intermediate occurs with proton transfer, via the active site water, from the Cys-NH_3^+ group onto the leaving ammonia (NH_3) derived from the glutamine substrate. A water is then activated by the now neutral Cys1 amine to act as a nucleophile to attack the thioester's carbonyl carbon (C_{carb}) to form a second covalently cross-linked tetrahedral oxyanion intermediate. This oxyanion then collapses with cleavage of the thioester's C_{carb}–S_{Cys1} bond to give glutamic acid (see Scheme 2).

Unfortunately, however, the exact details of the mechanism remain unclear. For instance, covalent modifications can impact catalytic efficiency via disruption of the structure of the glutaminase domain due to possible steric clashes with neighboring residues.[20] Furthermore, while its glutaminase activity was indeed markedly reduced by such modifications, it was not completely eliminated. As a result, Teplyakov et al.[9] have suggested that an ammonium-thiolate ion pair may be the resting state. In addition, the identity of the mechanistic base that activates (deprotonates) the thiol of Cys1 is not well understood.

We have used molecular dynamics (MD) simulations and quantum mechanical (QM)-cluster based computational approaches to investigate the protonation state of key active site residues and functional groups, e.g., the **α** -amine of Cys1. In addition, the influence of the protonation state of $_{Cys1}NH_2$ on the structure of the active site has been considered. Then, using an ONIOM (QM/MM) based approach the catalytic mechanism of the glutaminase domain of Glucosamine-6-phosphate synthase (GlmS) has been elucidated.

4.2 Computational Methods

4.2.1 Preparation of the Sample.

The X-ray crystal structure of a homodimer of glucosamine-6-phosphate synthase (GlmS) from Escherichia coli (Protein Data Bank entry 2J6H, 2.35 Å resolution) was

taken for the computational studies.[6] Mutation of the covalently-linked 5-oxo-L-Norleucine to glutamine in the glutaminase domain, and glucose-6-phosphate to fructose-6-phosphate in the synthase domain, were computationally applied.

The system was hydrogenated in accordance with the protonation states of various charged side chains with PROPKA.[17-20] It should be noted that histidyl were protonated according to their polar environment. Topology and coordinate files were built with the tleap module of AMBER14.[21] The system was solvated with 87921 TIP3P[22] water molecules, resulting in cubic boxes with an edge length of ~141.1 Å. The solvation achieved a density of 1.05 g/cm^3 after equilibration, and contained a total of 282727 atoms. It is worth mentioning that the solvated protein concentration was approximately one half of the original GlmS crystal structure.

4.2.2 Molecular Dynamics Simulations

The simulations were performed using the AMBER14 program software[21] with the cuda-enabled graphics processing units (GPUs) version of pmemd.[23,24] All the simulations used the ff14SB[25] and TIP3P[22] force fields for protein and waters, respectively. The neutral N-terminal cysteinyl, glutamine, and frucutose-6-phosphate (F6P) ligands were built with the antechamber[26,27] program using the ff14SB force field and RESP atomic charges. It should be noted that the charges of Cys1 and F6P fragments, capped by methyl groups, were derived from gas phase optimizations at the HF/6-31G(d) level of theory using Gaussian09 program.[28] Periodic boundary conditions, and an NVT ensemble were applied.[29] A cut-off of 8 Å was applied in real space for long-range electrostatic interactions using the Particle-Mesh Ewald (PME) procedure. A timestep of 2 fs was used for both equilibration and production by restricting bond stretches associated with hydrogen atoms through the SHAKE algorithm. The equilibrations were conducted in five stages after energy minimization: (1) Proper geometry of the hydrogen

atoms, all heavy atoms, including water oxygen, were restrained with a harmonic constant of 50 kcal mol^{-1} $Å^{-2}$ for 100.0 ps, at 10 K; (2) An identical potential and conditions were applied for an additional 100.0 ps, but with the removal of the restraint on the water oxygens to ensure optimized positions of water with respect to the protein environment; (3) These harmonic constants on the protein heavy atoms was decreased to 5 kcal mol^{-1} $Å^{-2}$ for 100.0 ps; (4) The harmonic potential was removed for 100.0 ps; (5) Finally, the system was gradually heated to 300 K over a time period of 2000.0 ps. The velocities were randomly updated every 10 steps for equilibration stages 1-4 and every 100 steps for stage 5. The production run was conducted for 150 ns following equilibration.

4.2.3 ONIOM (QM/MM) Calculations

All QM/MM calculations were performed within the ONIOM formalism as implemented in the Gaussian 09 suite of programs.[28] A suitable, representative structure was extracted from the MD simulation based on the most frequent RMSD population of Cα atoms in GlmS. Due to our interest in studying the deamination reaction of GlmS, our model was extended by two layers of surrounding residues from the deamination active site, and included 1512 atoms. For the mechanism involving N-terminal amino group acting as the base initiator, the QM region included Cys1, N and Cα atom (with corresponding hydrogens) of Gly2, all of Asn98 (except its amino group), N and Cα atom of Gln-99, and the glutamine substrate (apart from its amino group and carboxylic acid group). All other atoms were placed in the molecular mechanics (MM) layer, with regions on the exterior being restrained at their Cα atoms to keep the native conformation of the protein. Optimization, frequency, and Gibbs free energy corrections of structures were performed at the ONIOM [M062x/6-31G(d,p):AMBER96] level of theory within the mechanical embedding (ME) formalism.[30-32] The reason for the choice of the M062X functional was due to its ability to reproduce experimental kinetic values as validated by

several independent researchers in the field.[33-36] Single point calculations of the optimized structures were performed at the ONIOM [M062X/6-311++G(2d,p):AMBER96] ME level of theory, with the appropriate Gibbs free energy corrections.

4.3 Results and Discussion

4.3.1 Stability of the Simulations.

The root-mean-square deviation (RMSD) of the Cα atoms of residues within 20 Å surrounding the deamination site was monitored for GlmS with: (1) protonated and (2) neutral N-terminal amino group (Scheme 4.1). Along the course of the simulation, the RMSD value of the former stayed within the range of approximately 3.5-4.3 Å. The RMSD of the latter varied slightly between 0.8-1.2 Å with respect to the reference crystal structure. Protons cannot be resolved using conventional X-ray crystallography. However, the structure of GlmS with the neutral N-terminal Cys1 amine resembled its crystal structure, which suggests the protonation state of Cys1 is indeed neutral. However, in order to further investigate and confirm our hypothesis, reaction mechanisms of both possibilities were conducted using QM/MM methodologies and proton affinity calculations were conducted with QM-Cluster calculations.

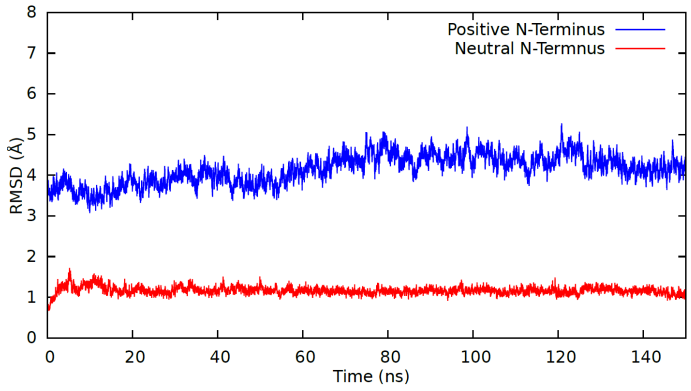


Figure 4.1. The RMSD fluctuations of GlmS of different protonation states. The blue denotes GlmS with positively charged N-terminus, while the red denotes GlmS with neutral N-terminus. (See text for more details.)

4.3.2 Interactions Surrounding the Glutamine Substrate.

Visual inspection of the GlmS, with protonated N-terminus, showed that the ligand changed its conformation dramatically with respect to the reference crystal structure. During the simulation, the $-NH_2$ side group of glutamine was found to form a hydrogen bond (H-bond) with the amino group of the glutamine ligand (Figure 4.2). Additionally, hydrogen bond donors such as Asn98 and Gly99 moved away significantly. The amide carbon centre of the ligand was also found to move away from the Cys1 thiol group to a distance of approximately 5.77 Å. This distance is too far for nucleophilic attack.

In GlmS with a neutral N-terminus, on the contrary, the ligand maintained interactions found in the crystal structure, such as its interactions with H-bonding donors (eg. Asn98 and Gly99). The distance of the thiol sulphur to the amide carbon centre was also found to be closer at approximately 3.35 Å. These results suggest that the positive charge in N-terminus disrupts the native interactions of the enzyme.

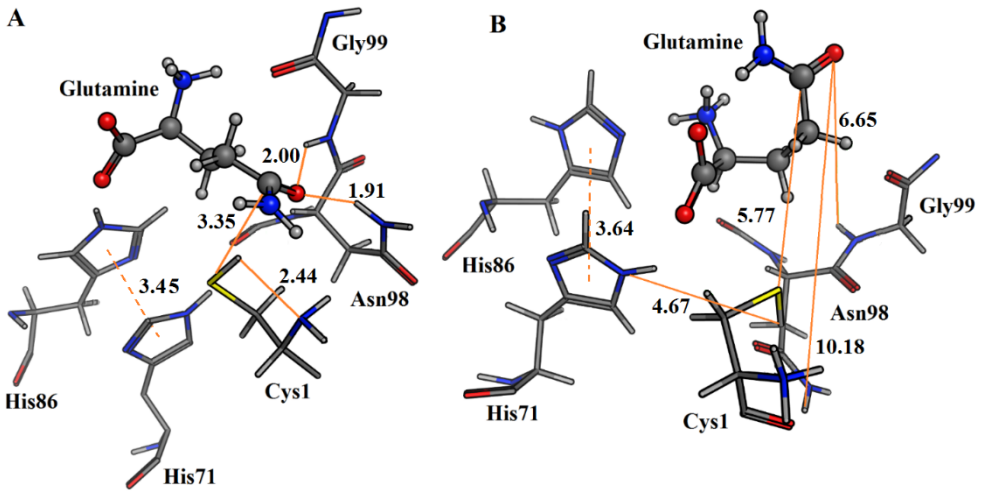


Figure 4.2. Representative structures, with selected distances shown in Angstroms, extracted from the MD simulations of the glutaminase domain of GlmS with a (A) neutral N-terminus, and a (B) protonated N-terminal Cys1 amine group.

4.3.3 Protonation States of Important Side Groups

GlmS is known to use its Cys1 thiolate side group to attack the amide carbon centre of the glutamine ligand. For this to occur, the thiol group must first be activated by a base. Two likely candidates exist for this: (1) the N-terminal amine group of Cys1 and (2) a nearby His71 residue. PROPKA[19] results have suggested that the N-terminus of GlmS is neutral with a very low pKa value of approximately 3.5. To validate the protonation states of various residues, a proton affinity study was employed (see computation methods). Our results indicate that the proton affinity of Cys1 thiol group is lower than methylthiol by 28.7 kJ mol^{-1}, and could be more easily deprotonated (Figure 4.3). His71 imidazole group showed a modest increase in proton affinity of 14.2 kJ mol^{-1}. Interestingly, the proton affinity of Cys1 amine group decrease dramatically from methylamine by 54.5 kJ mol^{-1}.

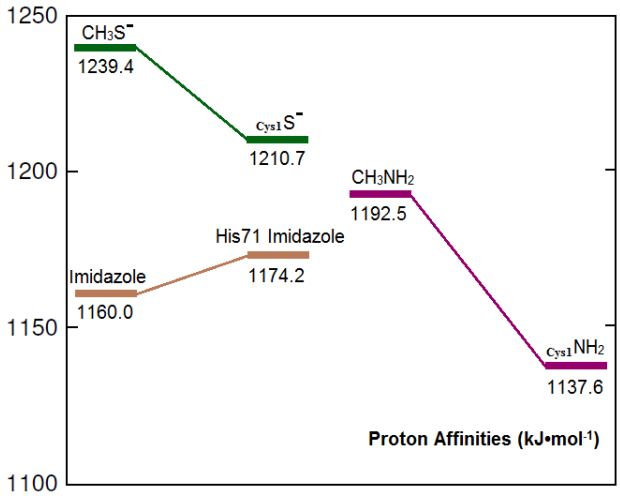


Figure 4.3. The calculated proton affinities (in kJ mol^{-1}) of various protein side chains compared to their simple molecular counterparts (see computational methods). The calculated proton affinity of water is 1010.0 in kJ mol^{-1} for reference.

Coupled with the MD simulation, these results suggest that the N-terminal amine is most likely neutral. However, because the proton affinity of N-terminal amine is quite poor, this raises the question of whether it could act as a base to abstract the thiol proton. A neutral His71 on the other hand would be a much better nucleophile, based on its proton affinity. Consequently, both reaction mechanisms (1) and (2) were explored using a QM/MM approach.

4.3.4 N-terminal Amine Acting as the Initial Base (Reaction 1)

In the experimentally proposed mechanism, the N-terminal amino group of Cys1 abstracts its thiol group for nucleophilic attack. For this to occur, the N-terminal amino group and imidazole ring of His71 were kept neutral. The potential energy surface (PES) in Figure 4.4 was explored using a combined QM/MM approach (See computational methods) while the corresponding optimized structures with important distances are

shown in Figure 4.5. In the optimized structure of the reactant complex (RC); TS1; IC1; IC5; TS6; and product complex (PC), the sulfur of the thiol group formed a hydrogen bond (H-bond) with the imidazole proton of His71. This effectively stabilized the Cys1 thiolate, allowing it to be deprotonated easily for a nucleophilic attack which is consistent with our proton affinity study above (Figure 4.3). Interestingly in RC; IC1; IC5; and PC (Figure 4.5), the Cys1 thiol/thiolate formed H-bonding interaction with the N-terminal amino group. As a consequence, the transfer of proton within Cys1 from its thiol group to the N-terminal amino group occurred directly with an energetic barrier of 67.1 kJ mol^{-1}. This large energy barrier associated with the Cys1 thiol deprotonation could be attributed to the amino group (pKa = 8.18) being a poorer proton acceptor than the thiol group (pKa = 5.07).

In the IC1 structure, the Cδ-Sγ distance was 3.5 Å. This distance was decreased to 2.0 Å in IC2 as a result of the nucleophilic attack of the thiolate against the amide carbon, which is slightly further than a typical sulphur-carbon bond of ~1.8 Å. Consequently, the amide centre changed from sp^2 trigonal planar with ∠Cγ-Cδ-Nε of 114.3° (IC1) to sp^3 with ∠Cγ-Cδ-Nε of 109.2° (IC2). The Cδ-Oγ bond distance saw a change from double bond character of 1.2 Å to a single bond character of 1.3 Å; the latter of which is stronger than most C-O single bonds. The formation of an oxyanion in IC2 was stabilized by three hydrogen bond donors, including the $–NH_2$ side group of Asn98, the –NH– backbone of Gly99, and a water in the active site. In IC2, the $–NH_2$ group of glutamine ligand participated in H-bond with the N-terminal amino group at a distance of 2.6 Å between the heavy atoms (Figure 4.5). Interestingly, an elongated distance of 1.3 Å between Cys1 amino nitrogen and proton was indicative that the latter was prone to transfer to the $–NH_2$ group of glutamine. Subsequently, a 7-membered ring was formed in TS3. IC3 resulted in

the thioester substrate intermediate and a free ammonia which was the lowest energy structure along the PES at -3.8 kJ mol^{-1}. In IC3, the Cδ-Sγ bond was strengthened and decreased to 1.8 Å while the Cδ carbon centre regained its sp^2 character.

The structures from RC to IC3 represents the first half of the deamination reaction. It is known from previous studies of GlmS and other amidotransferases, that the free NH_3 exits the glutaminase domain through an ammonia channel to the synthase site.[9] Consequently, we replaced the ammonia with a water, and overlaid the energy of IC3 with IC3' on the PES (Figure 4.4). The addition of water to the thioester intermediate represented the largest energetic barrier along the PES, which was also the rate limiting step of the reaction at 74.4 kJ mol^{-1} (Figure 4.5). Like TS3, TS4 also formed a 7-membered ring in which the oxygen of the water attacked the thioester carbon centre. Again in IC4, the Cδ carbon centre of the ligand changed from sp^2 to sp^3. This created the oxyanion which was stabilized by Asn98, Gly99, and a water. The Cδ-Sγ bond was elongated from 1.8 Å in IC3 to 2.0 Å in IC4. This weakened Cδ-Sγ bond allowed the thiolate to become a better leaving group *via* TS5, which was also a barrierless transition state. The removal of the thiolate produced the product of the reaction: glutamic acid. Finally, the regeneration of the thiol group occurred directly from Cys1 N-terminal amino group. Overall, the final PC lies lower in energy than the RC by -26.3 kJ mol^{-1}, indicating that the overall mechanism is exothermic.

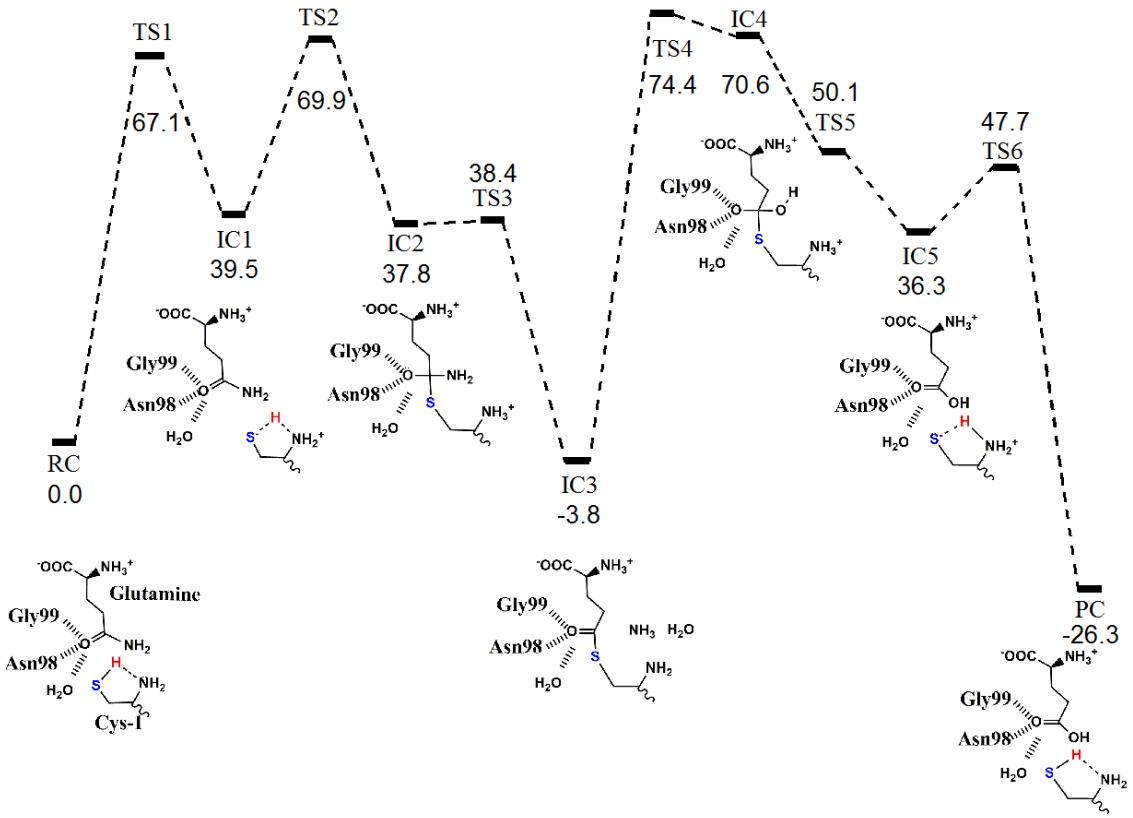


Figure 4.4 The free energy pathway (in kJ mol^{-1}) obtained for the conversion of glutamine to glutamic acid with release of ammonia wherein the N-terminal Cys1 amine of GlmS acts as the mechanistic base.

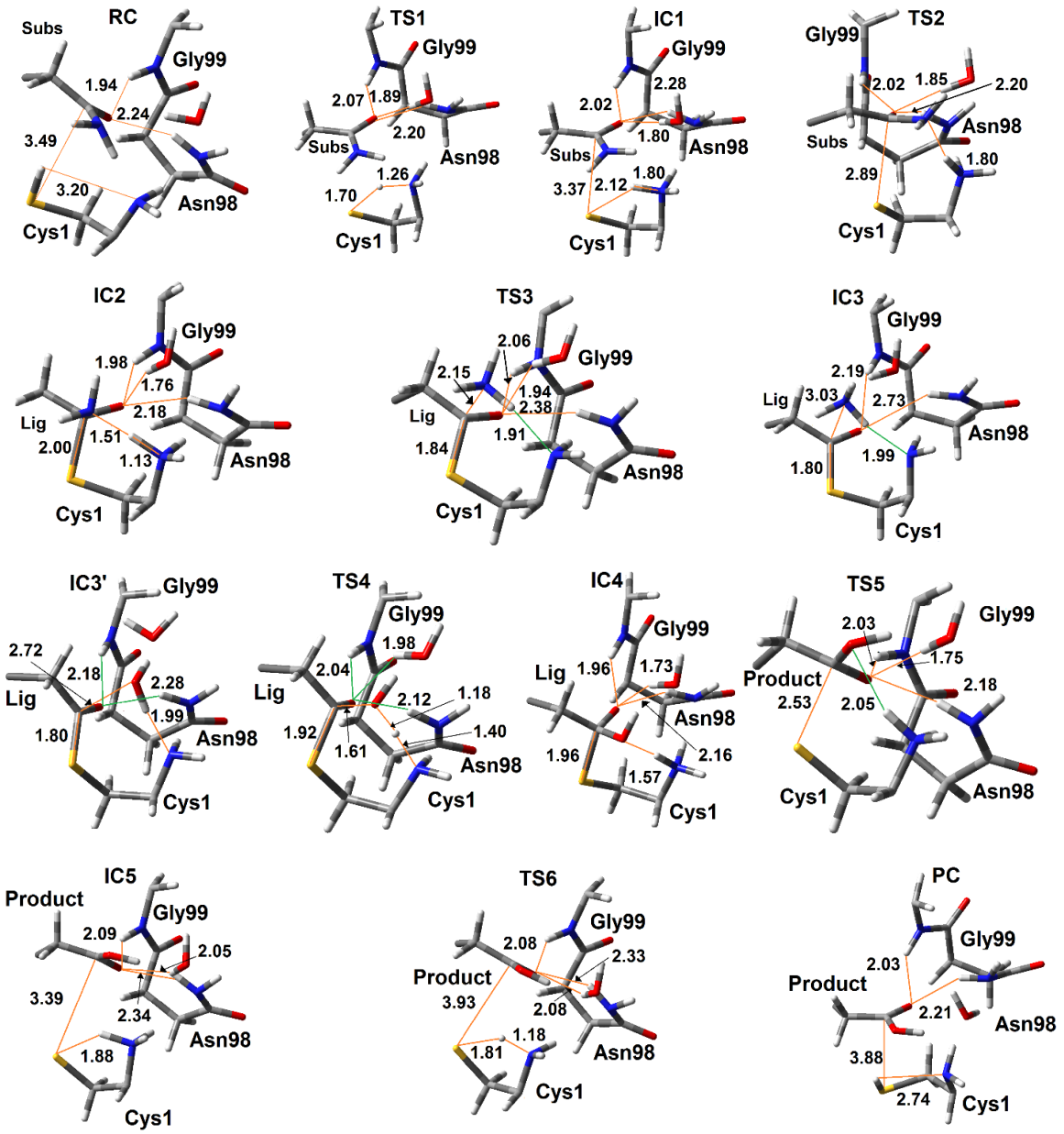


Figure 4.5. QM/MM optimized geometries (see Computational Methods) of intermediates, transition structures, and product complex obtained for the deamination mechanism of glutamine in which the N-terminal amine of Cys1 is neutral and acts as the catalytic base.

4.3.5 His71 Imidazole Group Acting as the Initial Base (Reaction 2)

Upon visual inspection of the active site, His71 was in close proximity to the Cys1 thiol group, and was a better nucleophile than the N-terminal amine according to our proton affinity studies (Figure 4.3). We took the RC of reaction 1, and protonated the N-terminus and rotated the dihedral bond corresponding to the Cα-Cβ-Cγ-Nδ by 180°. This placed the histidine within hydrogen bonding distance of the thiol hydrogen.

Unsurprisingly, the proton transfer from Cys1 thiol group to the His71 imidazole group is efficient via TS1 with a small energetic barrier of 4.1 kJ mol^{-1} (Figure 4.6). Structurally, IC1 to IC5 of reaction 2 was very similar to that of reaction 1 (Figure S10). Additionally, the RLS of both reaction pathways was the addition of the water to the thioester intermediate. The energy of this step was higher for reaction 2 than reaction 1 by approximately 20 kJ mol^{-1}, which further supports the fact that Cys1 N-terminal amine is likely deprotonated, and abstracts the proton from the Cys1 thiol group.

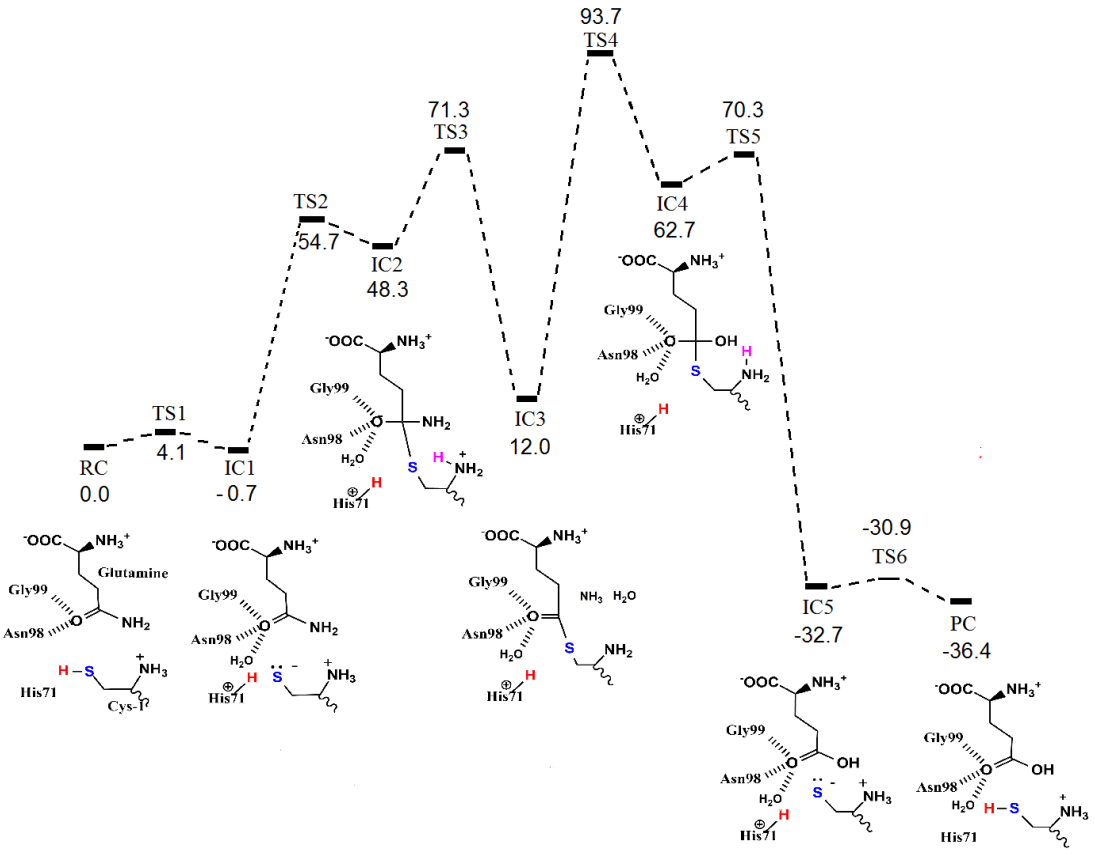


Figure 4.6. The free energy pathway obtained for the conversion of glutamine to glutamic acid with release of ammonia wherein the imidazole of His71 acts as the mechanistic base.

Interestingly, there were many energetic differences between these two pathways. For example in reaction 2, the energy of the thioester intermediates (IC3) was 12.0 kJ mol^{-1} above the RC but for reaction 1, it was 3.8 kJ mol^{-1} below. Likewise, the deamination step via TS3 was significantly higher in reaction 2 than reaction 1. These differences may be caused by the effect of the charged His71 residue in the active site. These QM/MM results, coupled with our MD and proton affinity studies suggest that a neutral Cys1 N-terminal amine group is most probable.

4.4 Conclusion

In this present study, multiscale computational methods have been used in a stepwise fashion to explore the glutaminase domain of the physiologically important GlmS enzyme, which is conserved in Class II Amidotransferases. Specifically, molecular dynamics (MD), combined quantum mechanics mol^{-1}ecular mechanics (QM/MM), and proton affinity studies were employed in synergy to elucidate the reaction mechanism of GlmS.

Cys1 thiolate was found to have depressed proton affinity compared to methylthiolate, which indicated that it could be deprotonated easily. Additionally, Cys1 N-terminal amine group is certainly neutral, and could abstract the proton directly from Cys1 thiol group. Subsequently, the Cys1 thiolate attacks the amide carbon centre of glutamine, and generate an oxyanion species. During the deamination of glutamine, the stabilization of several oxyanions intermediates was required. This was achieved through hydrogen bonding via the $–NH_2$ side group of Asn98, the –NH– backbone of Gly99, and a water. These critical H-bond donors were maintained for GlmS with a neutral N-terminus. On the contrary, GlmS with protonated N-terminus disrupted critical interactions, and effectively destabilized the active site.

The release of the ligand NH_2 group occurred simultaneously with the extraction of a proton from the N-terminal amino group via a 7-membered ring transition structure. The thioester intermediate following it was lower in energy than the reactive complex by 3.8 kJ mol^{-1}. The next step was the rate limiting step of the reaction with a barrier of 74.4 kJ mol^{-1}. The nucleophilic attack of water against the thioester carbon centre occurred also via a 7-membered ring structure. This regenerated the proton of the N-terminal amino group and formed yet another oxyanion on the ligand. After this step, the formation of the thiolate occurred, followed by rapid regeneration of the Cys1 thiol through abstraction of the proton from the N-terminal amino group. Overall, the process is exothermic, with a change in energy of -26.3 kJ mol^{-1}.

Although an alternative reaction mechanism whereby a nearby His71 extracts the proton of the Cys1 thiol group was deemed enzymatically feasible, it seems unlikely based on the protonation states of GlmS.

Overall, this study complements previous experimental evidences that the N-terminus of GlmS is neutral, and could act as a nucleophile for activating Cys1 thiol group. More importantly, we have been able to computationally elucidate GlmS deamination mechanism. In addition, we were able to identify key residues and characteristics of the reaction such as the rate limiting step.

4.5 Associated Content

Supporting Information.

A QM/MM potential energy surface the mechanism of His71 acting as a base. The supporting information is available in Appendix B.

4.6 References

(1) Massiere, F.; Badet-Denisot, M. A. The mechanism of glutamine-dependent amidotransferases. *Cell. Mol. Life Sci.* **1998**, *54*, 205-222.

(2) Boer, P.; Sperling, O. Role of Cellular Ribose-5-Phosphate Content in the Regulation of 5-Phosphoribosyl-1-Pyrophosphate and De-Novo Purine Synthesis in a Human Hepatoma-Cell Line. *Metabolism-Clinical and Experimental* **1995**, *44*, 1469-1474.

(3) Palmer, E. E.; Hayner, J.; Sachdev, R.; Cardamone, M.; Kandula, T.; Morris, P.; Dias, K. R.; Tao, J.; Miller, D.; Zhu, Y.; Macintosh, R.; Dinger, M. E.; Cowley, M. J.; Buckley, M. F.; Roscioli, T.; Bye, A.; Kilberg, M. S.; Kirk, E. P. Asparagine Synthetase

Deficiency causes reduced proliferation of cells under conditions of limited asparagine. *Mol. Genet. Metab.* **2015**, *116*, 178-186.

(4) Vanoni, M. A.; Curti, B. Structure-function studies of glutamate synthases: A class of self-regulated iron-sulfur flavoenzymes essential for nitrogen assimilation. *IUBMB Life* **2008**, *60*, 287-300.

(5) van den Heuvel, R. H. H.; Svergun, D. I.; Petoukhov, M. V.; Coda, A.; Curti, B.; Ravasio, S.; Vanoni, M. A.; Mattevi, A. The active conformation of glutamate synthase and its binding to ferredoxin. *J. Mol. Biol.* **2003**, *330*, 113-128.

(6) Mouilleron, S.; Badet-Denisot, M. A.; Golinelli-Pimpaneau, B. Glutamine binding opens the ammonia channel and activates glucosamine-6P synthase. *J. Biol. Chem.* **2006**, *281*, 4404-4412.

(7) Hebert, L. F., Jr.; Daniels, M. C.; Zhou, J.; Crook, E. D.; Turner, R. L.; Simmons, S. T.; Neidigh, J. L.; Zhu, J. S.; Baron, A. D.; McClain, D. A. Overexpression of glutamine:fructose-6-phosphate amidotransferase in transgenic mice leads to insulin resistance. *The Journal of Clinical Investigation*, *98*, 930-936.

(8) Teplyakov, A.; Obmolova, G.; Badet-Denisot, M. A.; Badet, B. The mechanism of sugar phosphate isomerization by glucosamine 6-phosphate synthase. *Protein Sci.* **1999**, *8*, 596-602.

(9) Teplyakov, A.; Leriche, C.; Obmolova, G.; Badet, B.; Badet-Denisot, M. A. From Lobry de Bruyn to enzyme-catalyzed ammonia channelling: molecular studies of D-glucosamine-6P synthase. *Nat. Prod. Rep.* **2002**, *19*, 60-69.

(10) Marshall, S.; Bacote, V.; Traxinger, R. R. Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. *J. Biol. Chem.* **1991**, *266*, 4706-4712.

(11) Milewski, S.; Janiak, A.; Wojciechowski, M. Structural analogues of reactive intermediates as inhibitors of glucosamine-6-phosphate synthase and phosphoglucose isomerase. *Arch. Biochem. Biophys.* **2006**, *450*, 39-49.

(12) Khan, M. A.; Gopel, Y.; Milewski, S.; Gorke, B. Two Small RNAs Conserved in Enterobacteriaceae Provide Intrinsic Resistance to Antibiotics Targeting the Cell Wall Biosynthesis Enzyme Glucosamine-6-Phosphate Synthase. *Front. Microbiol.* **2016**, *7*, 908.

(13) Bearne, S. L.; Blouin, C. Inhibition of Escherichia coli glucosamine-6-phosphate synthase by reactive intermediate analogues. The role of the 2-amino function in catalysis. *J. Biol. Chem.* **2000**, *275*, 135-140.

(14) Wojciechowski, M.; Milewski, S.; Mazerski, J.; Borowski, E. Glucosamine-6-phosphate synthase, a novel target for antifungal agents. Molecular modelling studies in drug design. *Acta Biochim. Pol.* **2005**, *52*, 647-653.

(15) Bearne, S. L. Active site-directed inactivation of Escherichia coli glucosamine-6-phosphate synthase. Determination of the fructose 6-phosphate binding constant using a carbohydrate-based inactivator. *J. Biol. Chem.* **1996**, *271*, 3052-3057.

(16) Mouilleron, S.; Badet-Denisot, M. A.; Badet, B.; Golinelli-Pimpaneau, B. Dynamics of glucosamine-6-phosphate synthase catalysis. *Arch. Biochem. Biophys.* **2011**, *505*, 1-12.

(17) Bas, D. C.; Rogers, D. M.; Jensen, J. H. Very fast prediction and rationalization of pK(a) values for protein-ligand complexes. *Proteins-Structure Function and Bioinformatics* **2008**, *73*, 765-783.

(18) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK(a) values. *Proteins-Structure Function and Bioinformatics* **2005**, *61*, 704-721.

(19) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525-537.

(20) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK(a) Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284-2295.

(21) D.A. Case, J. T. B., R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman. AMBER 2015. *University of California, San Francisco* **2015**.

(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

(23) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542-1555.

(24) Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878-3888.

(25) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

(26) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247-260.

(27) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

(28) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision D.01. *Gaussian, Inc., Wallingford CT* **2009**.

(29) Andersen, H. C. Molecular-Dynamics Simulations at Constant Pressure and-or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384-2393.

(30) Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372-1377.

(31) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668-1688.

(32) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

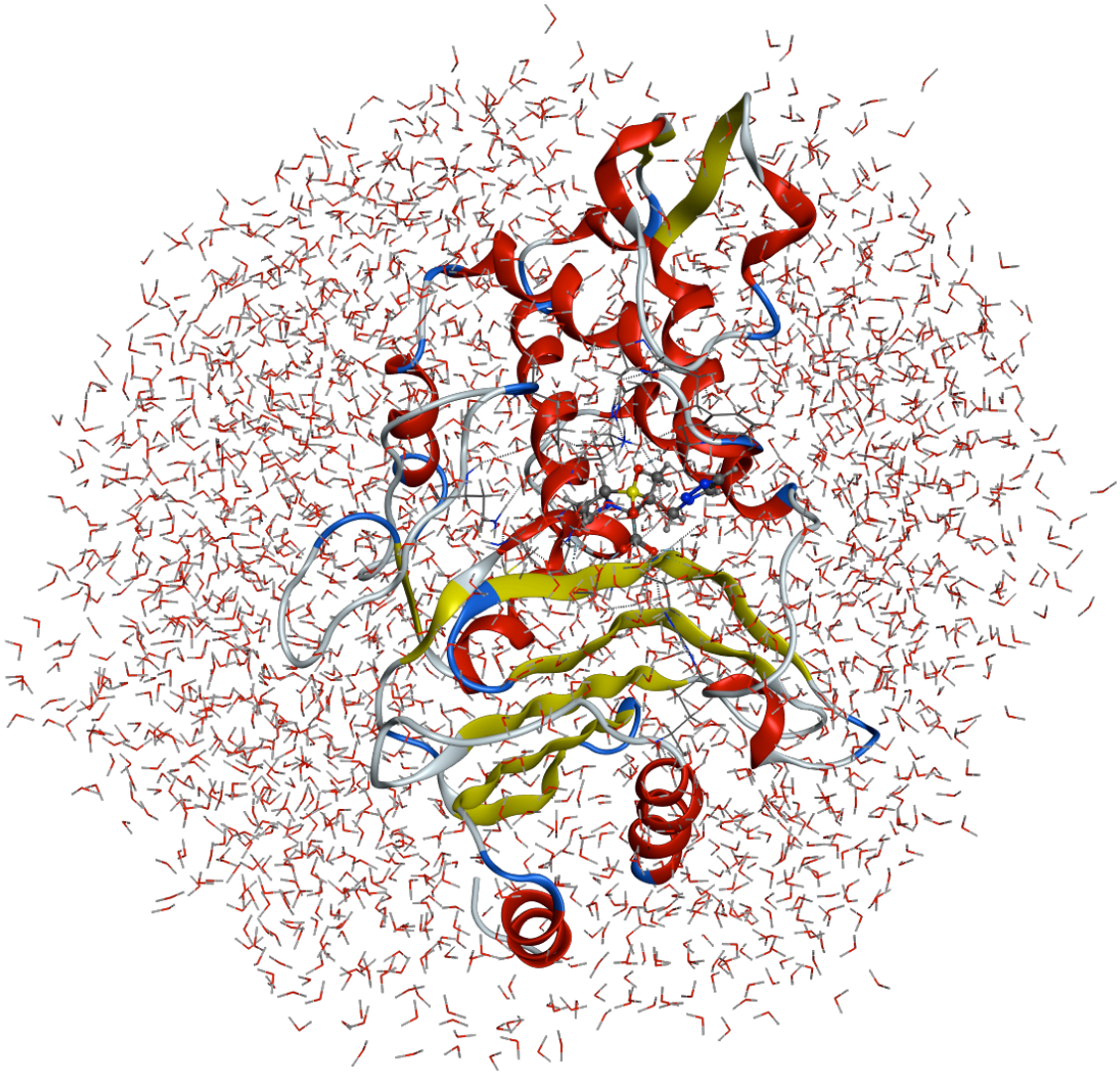
(33) Parkhomenko, D. A.; Edeleva, M. V.; Kiselev, V. G.; Bagryanskaya, E. G. pH-sensitive C-ON bond homolysis of alkoxyamines of imidazoline series: a theoretical study. *J. Phys. Chem. B* **2014**, *118*, 5542-5550.

(34) Kumar, M.; Busch, D. H.; Subramaniam, B.; Thompson, W. H. Role of tunable acid catalysis in decomposition of alpha-hydroxyalkyl hydroperoxides and mechanistic implications for tropospheric chemistry. *J. Phys. Chem. A* **2014**, *118*, 9701-9711.

(35) Alarcon, P.; Bohn, B.; Zetzsch, C.; Rayez, M. T.; Rayez, J. C. Reversible addition of the OH radical to p-cymene in the gas phase: multiple adduct formation. Part 2. *Phys. Chem. Chem. Phys.* **2014**, *16*, 17315-17326.

(36) Denegri, B.; Matic, M.; Kronja, O. A DFT-based model for calculating solvolytic reactivity. The nucleofugality of aliphatic carboxylates in terms of Nf parameters. *Org. Biomol. Chem.* **2014**, *12*, 5698-5709.

Chapter 5: Conclusion



5.1 Conclusions

In this thesis, multi-scale computational studies were used to study enzymes containing more than one catalytic active site. Specifically, the chemistry involved in the pre-transfer editing of yeast mitochondrial threonyl-tRNA synthetases (MST1) and the deamination of GlmS were elucidated.

In chapter 2, the reaction mechanism of pre-transfer editing of MST1 was investigated using microsecond molecular dynamics (MD) simulations and umbrella sampling, employing quantum mechanics/molecular mechanics simulations (QM/MM-MD). Additionally, the basis on how MST1 is able to discriminate between the cognate threonyl-AMP (Thr-AM) and the noncognate seryl-AMP (Ser-AMP) has been elucidated. Our results indicate that the aminoacyl-AMP (aa-AMP) uses substrate-assisted pre-transfer editing mechanism, whereby a catalytic water engages in the nucleophilic attack of its carbonyl centre. The catalytic water is activated through the phosphate oxygen of the aa-AMP via the extraction of its proton. This is a two-step process. The first step results in the formation of a tetrahedral carbon intermediate and an oxyanion. Interestingly, the oxyanion in Thr-AMP was stabilized by less hydrogen bonding waters than Ser-AMP. As a result, Ser-AMP is preferentially hydrolyzed due to kinetic and thermodynamic stabilization. Overall, our MD simulations indicate that the greater water density is caused by the decreased steric interactions of Ser-AMP compared to Thr-AMP, which increased water density in the active site of the former.

In the second and last step, the formation of the amino acid and adenosine monophosphate (AMP) occurs through the elongation of the carbon-phosphate oxygen bond, which changed the sp^3 character of the carbon centre to sp^2. These results are in agreement with experimental studies.[1,2]

In chapter 4, the deamination active site of GlmS was explored using MD simulations and ONIOM methods. Existing literature predicted that the N-terminus of GlmS is neutral,

and required for activating the Cys1 thiol group.[3] We were able to confirm their predictions through MD simulations. Interestingly, our results showed that a neutral N-terminus is crucial for the structural integrity of the enzyme, and has a favourable free energy barrier compared to an alternative mechanism. Specifically, we were able to show that the rate limiting step of the reaction was the nucleophilic attack of water against the thioester carbon intermediate, which had a free energy barrier of 74.4 kJ mol^{-1}. The energy barrier of the alternative mechanism, involving a neutral His71 acting as a base, was higher at 93.7 kJ mol^{-1}. Several enzyme residues, including the side group of Asn98 and the amine hydrogen of Gly99, and a water were seen to stabilize a recurring oxyanion in both reaction mechanism. Overall, based on our MD, ONIOM and proton affinity study, the reaction mechanism whereby the N-terminus extracts the proton is favoured over His71 acting as a base.

MST1 and GlmS are just two different examples studied, and this thesis does not address all aspects of the chemistry of multi-active site enzymes. However, significant contributions have been made in both categories of biomolecules. For example, we discovered that MST1 uses a substrate assisted mechanism for pre-transfer editing, which could be extended to all threonyl-tRNA synthetases and perhaps other aminoacyl-tRNA synthetases as well. GlmS, on the other hand, shares its N-terminal deamination domain with other type II amidotransferases. This class of enzymes encompasses many other enzymes, involved in different physiological processes from bacterial cell wall synthesis to nucleotide biosynthesis to glutamic acid synthesis. By understanding the chemistry of these two enzymes, it could lead to improvements in the design of pharmaceutical agents against a plethora of diseases, which are associated with these enzymes. The field of computational enzymology is still relatively young, and there is still much left to do in future works. These include identifying novel chemistry within enzymes as well as designing pharmaceutical drugs using computational chemistry.

5.2 References

(1) Ling, J. Q.; Peterson, K. M.; Simonovic, I.; Soll, D.; Simonovic, M. The Mechanism of Pre-transfer Editing in Yeast Mitochondrial Threonyl-tRNA Synthetase. *J. Biol. Chem.* **2012**, *287*, 28518-28525.

(2) Ling, J.; Peterson, K. M.; Simonovic, I.; Cho, C.; Soll, D.; Simonovic, M. Yeast mitochondrial threonyl-tRNA synthetase recognizes tRNA isoacceptors by distinct mechanisms and promotes CUN codon reassignment. *Proc Natl Acad Sci U S A* **2012**, *109*, 3281-3286.

(3) Massiere, F.; Badet-Denisot, M. A. The mechanism of glutamine-dependent amidotransferases. *Cell Mol Life Sci* **1998**, *54*, 205-222.

Appendix A

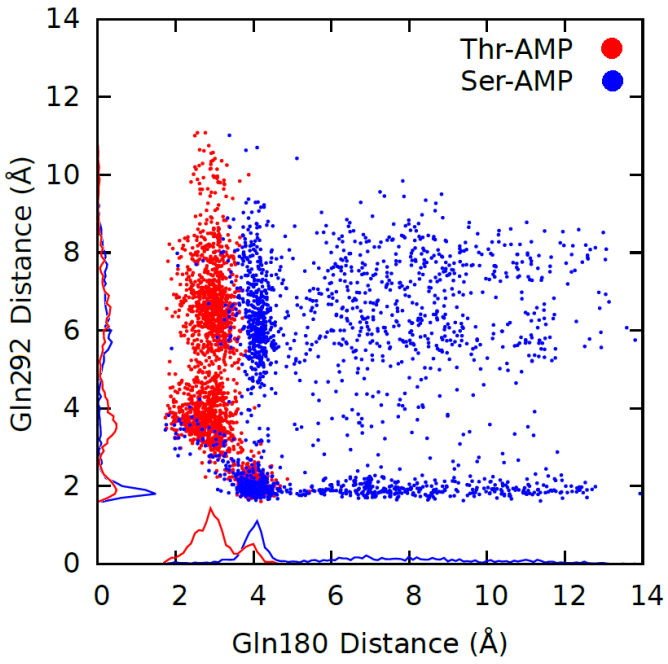


Figure S1. Thr-AMP (red) is able to bind MST1 in two distinct conformations by switching H-bonding to phosphate group between Gln180 (x-axis) and Gln292 (y-axis) through rotation of its phosphate group. Along each axes, a probability distribution of the distances are shown.

Benchmark Modeling of Ser-AMP Hydrolysis

We performed a static QM calculation in order to find a suitable semi-empirical method to use for QM/MM-MD umbrella sampling. This is because of the high computational costs associated with umbrella sampling employing Density Functional Theory (DFT). Geometry optimization were performed with Gaussian09 rev.D software[1] using DFT with B3LYP/6-31G**. The polar environment was provided by Conductor-Like Polarizable Continuum Model (CPCM, $\varepsilon = 78.5$).[2,3]. The reactant (RC), product (PC), and transition states (TS) were successfully optimized. In addition, an intrinsic reaction coordinate calculation was run on the transition in order to find good descriptors for the reaction. This information is shown below.

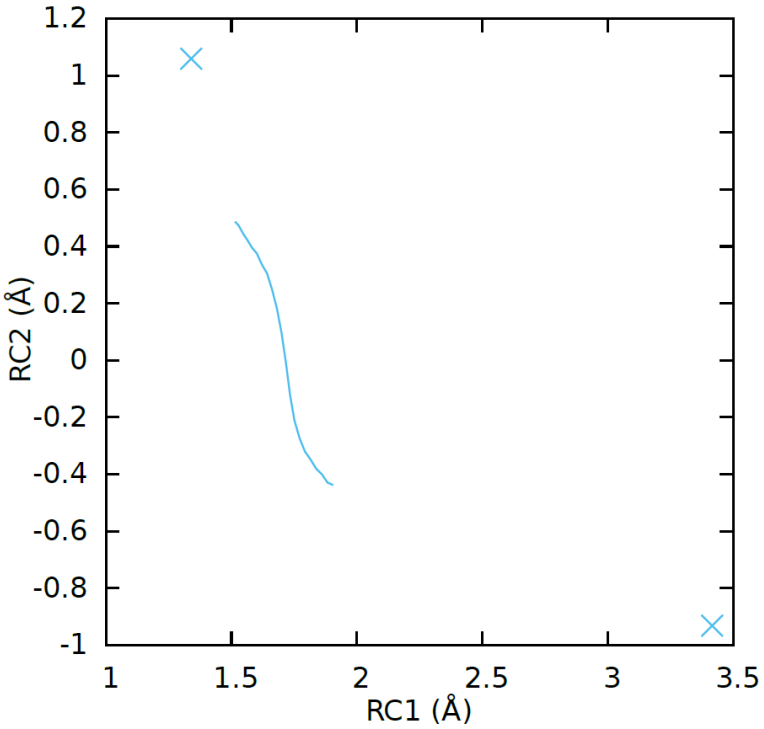


Figure S2. The reaction coordinates defined for the pre-transfer editing hydrolysis using CPCM B3LYP/6-31G(d,p).

Evaluation of Different Semi-empirical Methods

AM1/d

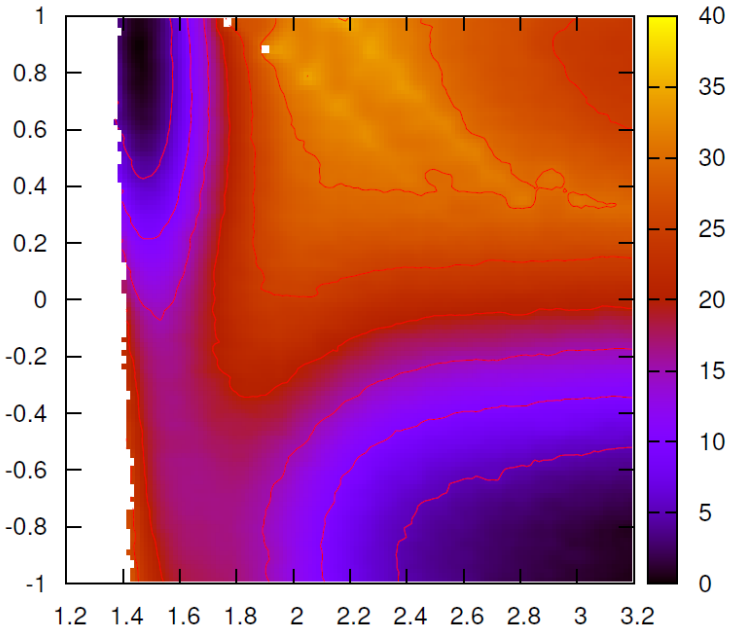


Figure S3. The potential energy surface generated by AM1/d method.

PM3

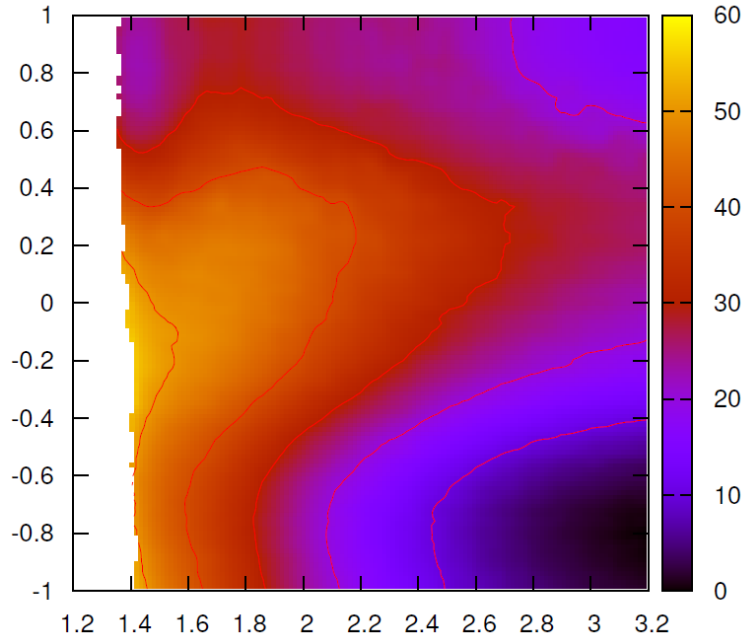


Figure S4. The potential energy surface generated by PM3 method.

PM6

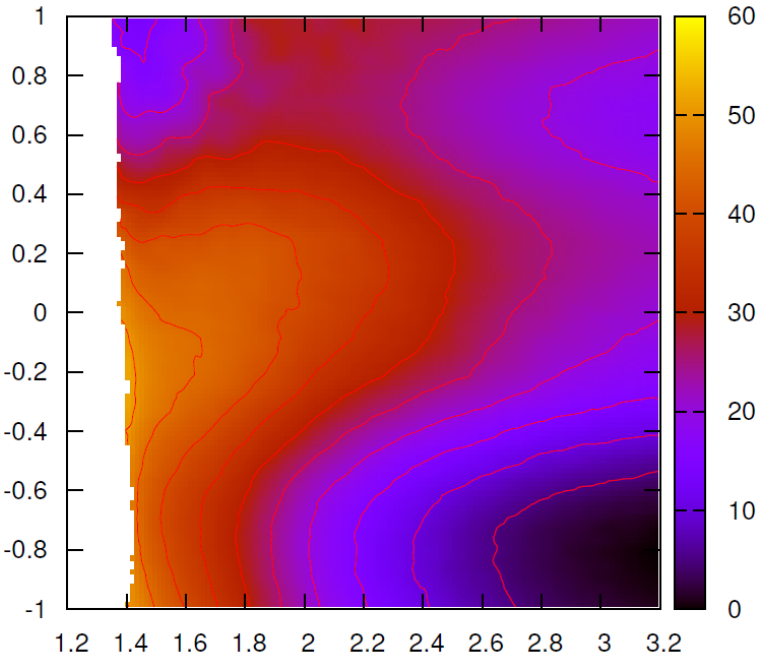


Figure S5. The potential energy surface generated by PM6 method.

AM1

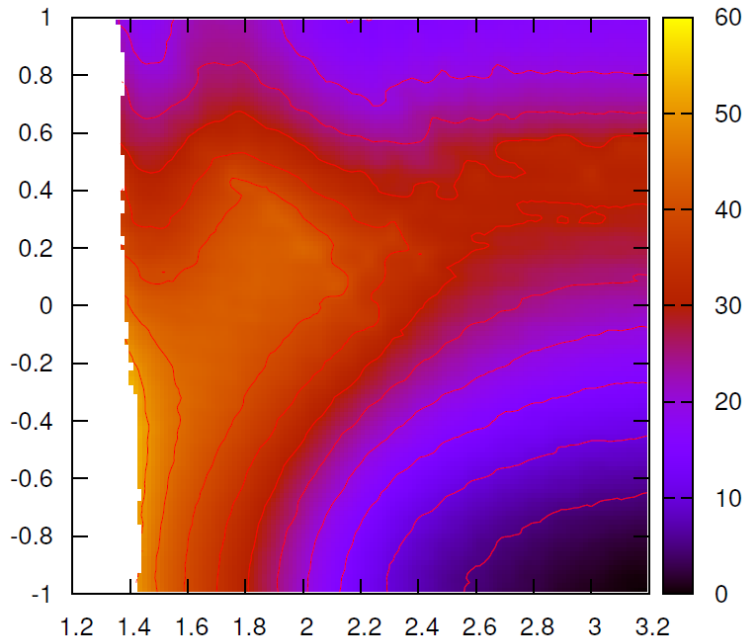


Figure S6. The potential energy surface generated by AM1 method.

Validating the Formation of the Product Complex. Since the second step of the reaction changes the reactant from the tetrahedral intermediate to the trigonal planar intermediate, it is possible to determine the successful formation of the product complex by observing the improper angles associated with the structure. Below, we have highlighted the definition of the improper angle used. Values closer to 0 signify the successful formation of the product complex.

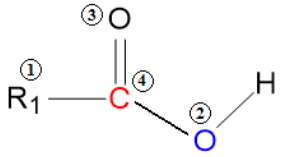


Figure S7. Definition of improper angle used to confirm the formation of the product complex from the intermediate complex.

For both Thr-AMP and Ser-AMP, the improper angles show that the product was indeed formed from the intermediate since there was a small deviation from 0 degrees (Figure S9) from 30 degrees.

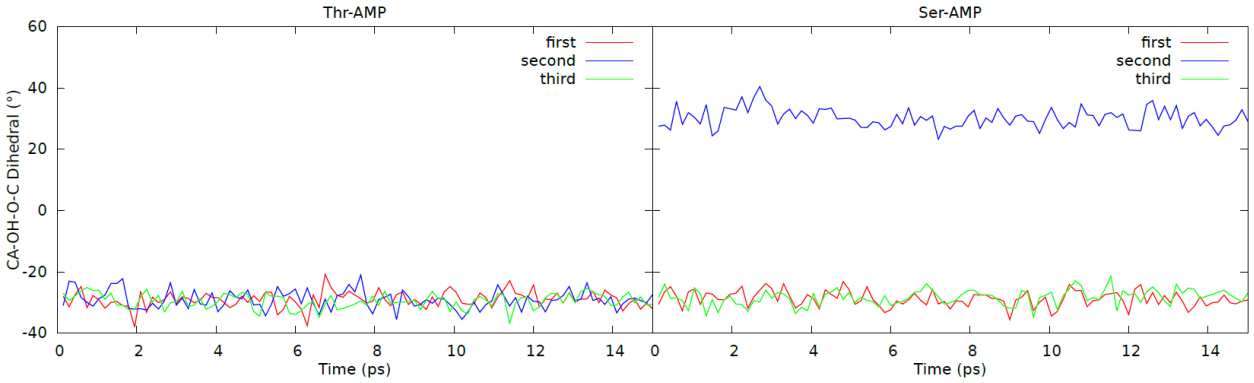


Figure S8. The improper angle defined in *Figure S7* over 15 ps in the sampling window corresponding to d_1= 1.5 Å for both Thr-AMP (left) and Ser-AMP (right).

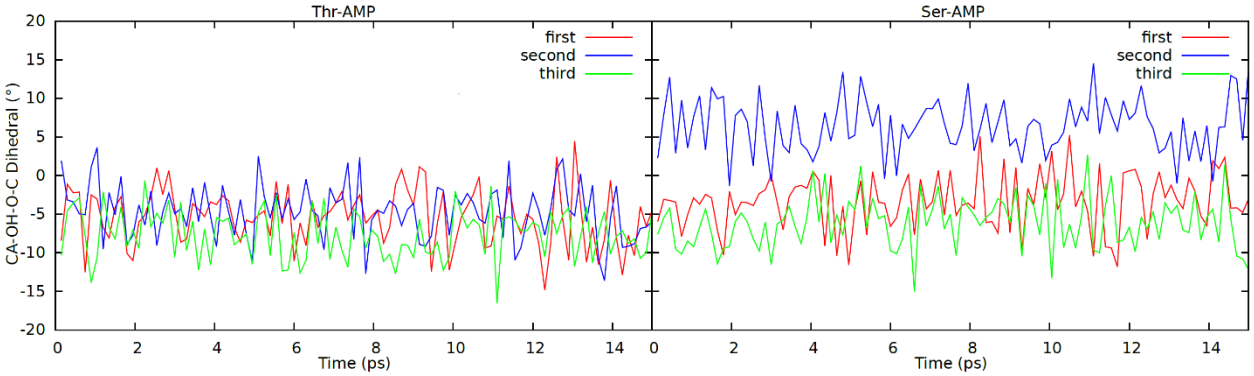


Figure S9. The improper angle defined in *Figure S7* over 15 ps in the sampling window corresponding to d_1= 2.4 Å for both Thr-AMP (left) and Ser-AMP (right).

References

(1) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision D.01. *Gaussian, Inc., Wallingford CT* **2009**.

(2) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J Comput Chem* **2003**, *24*, 669-681.

(3) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995-2001.

Appendix B

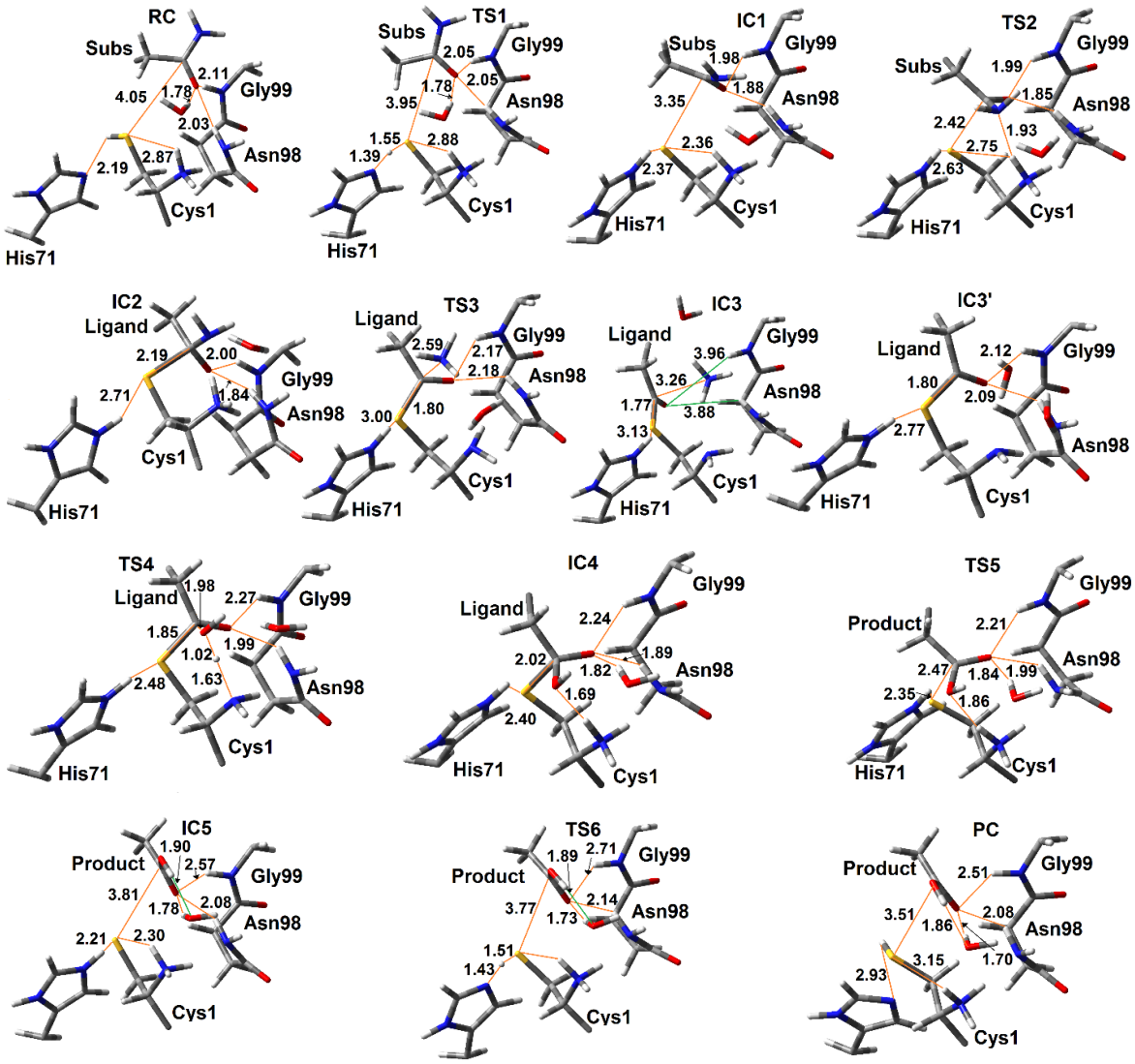


Figure S10. Optimized intermediates and transition state structures within the deamination site of the GlmS where the N-terminal amino group is positively charged, and a nearby His71 acts as a catalytic base.

Vita Auctoris

Name:	Wanlei Wei
Place of Birth:	Lanzhou, China
Year of Birth:	1991
Education:	Vincent Massey Seconary School, Windsor, ON 2009
	University of Toronto, B.Sc., Toronto, ON, 2014
	University of Windsor, M.Sc., Windsor, ON, 2016